Computational Aspects of Language Acquisition from World to Word

Area Paper

Frank A. Smadja

December 21, 1988

Cucs-399-88

Contents

List of Figures

1 Introduction

This paper surveys and evaluates research on the subject of language acquisition. Emphasis is put on computational aspects of the problem, and our discussion is based on a critical presentation of a group of systems that have been developed and that we found the most relevant to our topic.

By computational aspects of language acquisition, we mean the ability to parallel in a computer program human faculties of learning or acquiring a language. Our concern is the acquisition of linguistic knowledge. Besides syntax, linguistic knowledge can be separated into two distinct parts: knowledge relevant to the dictionary (or lexicon), and knowledge relevant to the encyclopedia [Clark 77]. The encyclopedia describes the world while the dictionary is more directed towards the internal relations of the lexicon. We use this distinction as a criteria on which to base our presentation of the programs described in this paper.

Building a model of language acquisition is a hard task. Most problems encountered are not restricted to language acquisition but are also relevant to the description of language generation and analysis. The programs we have chosen to present here, are representative of the main choices and commitments at the conceptual level as well as at the implementation level.

Language acquisition has been extensively studied in the fields of both linguistics and psycholinguistics. Research in linguistics has focused on the language aspect of the problem while research in psycholinguistics has put the emphasis on the psychological framework of the acquisition. Our discussion is based upon these two aspects of language acquisition and we review the systems from these two viewpoints.

The first two chapters are a critical presentation of five systems we have chosen to focus on, and the last two chapters are of more theoretical content. We have distinguished three main kinds of knowledge used by the programs involved here: syntactic, lexical and encyclopedic knowledge. In chapter two, for each kind of knowledge, we first describe the programs using it and then evaluate the use of this kind of knowledge in the field of computational linguistics in general and of language acquisition in particular.

Section three describes, compares and contrasts the types of learning strategies and biases used. During the presentation, we focus on the psychological and linguistic effects of the choices made, and we introduce the need for a sound approach to this problem.

The rest of the paper is dedicated to more theoretical work. Psycholinguistics and linguistic results in language acquisition are sketched in chapter four, and the main hypotheses and models of the young field of second language acquisition are presented as well.

Chapter five draws on the previous chapters and concentrates on lexical knowledge. Although lexical knowledge has been investigated in different research fields, there is a common tendency towards its formalization and systematic account in linguistic models. We present these investigations and evaluate them as the basis for a computational approach to language acquisition.

2 Goals, Assumptions and Knowledge Representation

In this section we review and discuss the programs in turn according to the kind of knowledge they use, whether knowledge of the encyclopedia, knowledge of the lexicon or syntactic knowledge. For each type, we give a brief overview of the programs using it, and we then describe and discuss the use of this kind of knowledge. We have chosen to focus on five programs all aiming to acquire language-related knowledge: FOUL-UP [Granger 77], CHILD [Selfridge 80], Berwick's program [Berwick 85], Debili's program [Debili 85], and RINA [Zernik 87]. We review them in turn and try to locate them amidst the related theoretical works in both fields of linguistics and language acquisition. They all acquire a subset of language-related knowledge, and they can be partitioned into three groups: the programs acquiring encyclopedic knowledge (RINA, FOUL-UP and CHILD), the programs acquiring syntax (Berwick's program) and the program acquiring lexical knowledge (Debili's program). We review the systems in turn according to this classification.

2.1 Programs Acquiring Encyclopedic Knowledge: FOUL-UP, CHILD and RINA

Notwithstanding their differences in terms of goals and techniques used, RINA, FOUL-UP and CHILD are based upon the same conception of natural language. This conception is derived from theories of conceptual dependency [Schank 72] and dynamic memory [Schank 82]. In this work, language is considered as a means to convey information about the actual world and, as such, the emphasis is put on its semantic aspects. In this section, we first describe the programs separately and then describe and evaluate from our perspective the use of conceptual dependency as a representation technique for the treatment of language acquisition.

2.1.1 Words' Meaning: FOUL-UP

FOUL-UP was designed in 1977 by Richard Granger at Yale University [Granger 77] and is one of the first programs to attempt to acquire language. Its task is to figure out the meaning of words from context. FOUL-UP is provided with a knowledge base of mundane events stored as scripts. FOUL-UP parses English sentences and, when encountering an unknown word, looks up the script based memory and attempts to understand the unknown word in a particular context. The contexts or situations are encoded in the script-based memory, and are triggered according to the sentence. In this case, the context is limited to what has been termed the semantic situation [Morris 71], *i.e.* the world referent of the sentence. In the program, the world context is represented in the form of several chunks of knowledge, and the unknown word is understood in light of the situation encountered. The acquired knowledge is not purely "linguistic" but clearly encyclopedic.

If provided with the following sentence:

"Friday, a car swerved off Route 69. The car struck an elm."

FOUL-UP knows all the words in the sentence except *elm*. The semantic situation referred to in this sentence is the vehicle-accident situation. In the program, it is represented by a script (the vehicle-accident script) containing typical information about the way accidents occur. Using this vehicle-accidents script FOUL-UP is able to deduce part of the meaning of *elm*. The deduced properties for *elm* are that, it is a noun, a physical object, and plays the role of an obstruction in a car accident. This is not the full meaning of the word *elm*, but it captures its purport in some specific contexts.

FOUL-UP is an integral part of the SAM system [Cullingford 78] and aims at giving it more flexibility. As such it is built upon the basic assumptions of conceptual analysis. In addition to FOUL-UP, SAM is composed of three main interacting modules: ELI, the integrated parser [Riesbeck 76], *TOK*, the tokenizer, and APPLY, the script applier. Each sentence is first parsed[1] by ELI which returns its conceptual dependency representation. Then *TOK* completes the representation in order to simplify the treatment of anaphoric reference and disambiguation, and finally APPLY maps between the representation and the script based memory of world events in triggering the relevant script. The learning mechanism of FOUL-UP in this environment is simple: when an unknown word is encountered by ELI during the parsing process, FOUL-UP is called and takes note of the current expectations from the semantic situation and uses a "place-holder" to hold the information retrieved from the conceptual dependency representation of the sentence. This place-holder is used as a blackboard where each module will in turn deposit and withdraw some information until the actual end of the parse. For example, ELI will deduce syntactic features about the unknown while APPLY will deduce its world referent. Ultimately, this place-holder will represent the conceptual structure of the word. There is very little available description of FOUL-UP, although Granger claims to account for cognitive processes of first language acquisition in his work. Very little description of FOUL-UP is available and very few results have been given. No evaluation of the program has been found in the literature.

2.1.2 Child Language Acquisition: CHILD

CHILD [Selfridge 80] is a computer model of first language acquisition (*FLA*). Selfridge's program is a model of child comprehension and synthesis abilities between age 1 and age 2. His work and thesis pivot around long hours of observation of a group of children's linguistic behavior. The techniques and representation methods for the treatment of linguistic knowledge used in CHILD are also based upon Schank's work. However, whereas other programs described in this paper make implicit use of psychological knowledge, CHILD makes explicit use of it. The acquisition of language is considered as being part of more general learning activities. Psychological (or behavioral) knowledge is encoded in a simple rule based inference engine. For example, child smiling activities or attention capacities are simulated with simple inference rules. Figure 1 gives an example of a rule accounting for part of CHILD learning abilities.

[1] We presume that there is only one parse.

```
+-----------------------------------------------+
|          Gestural Meaning Inference           |
|                                               |
| If an utterance containing an unknown word is|
| spoken with a gesture having a clear meaning,|
| then hypothesize that the meaning of the word|
| is the same as the meaning of the gesture.    |
|                                               |
+-----------------------------------------------+
```

Figure 1: An example behavior rule for CHILD, from [Selfridge 82]

We give in figure 2 a simple example session of CHILD reactions at a stage when CHILD has already learned what *ball* means. This session is taken directly from Selfridge's thesis who edited the original computer output. Selfridge doesn't give much information on what happens between the user input sentence, (*"give me the ball"*) and CHILD's output, (ATRANS ACTOR (CHILD) OBJECT BALL). "STARES BLANKLY AT PARENT" and "RETURNS TO PLAY" are interpretations of the program's behavior. As for CHILD's visual information such as, the user holding out his/her hand or looking at the ball, no indication is given on the way to transmit it to CHILD. It is probably user supplied in a particular format which is not specified.

CHILD uses its knowledge of what can be done with a ball combined with world and semantic knowledge to infer what the parent wanted him to do.

Selfridge's work is focused more on the psychological modeling of $L1$[2] acquisition than on its linguistic counterpart. Selfridge considers that a language is learned by applying general learning mechanisms to linguistic data. His focus was on the modeling of those mechanisms in the context of language acquisition. In this respect, his work is more relevant to cognitive modeling than to computational linguistics as language per se is not the main concern here. This explains why most pure linguistic knowledge is not accounted for in his work.

From a psychological standpoint, children's behavior is not satisfactorily modeled by the 15 inference rules used by CHILD. Psychology of young children is intricate. Aspects of child psychology such as cognitive development, linguistic and affective development are difficult to model in a computer program as no psychological theory provides a complete and formal model of children's reactions. All aspects of child development are intertwined and difficult or impossible to simulate separately. The program heavily relies on arbitrary choices and simplifications, which I believe, weaken the results.

[2] In the rest of this paper, we will call $L1$ the first language (or mother tongue), and $L2$ a second language. Thus following the notation used in second language research works.

```
Parent (or User): give me the ball
CHILD stares blankly at parent
CHILD returns to play
CHILD sees: (parent holds out hand)
            (parent looks at the ball)
CHILD responds:
(ATRANS ACTOR (CHILD) OBJECT (BALL)
        TO (POSS VAL (PARENT)))
```

Figure 2: A sample session of CHILD, from [Selfridge 80]

2.1.3 Learning Idioms: RINA.

Zernik in [Zernik 87] describes RINA, a program based upon a dynamic hierarchical phrasal lexicon ($DHPL$) modeling $L2$ acquisition of English idioms and phrases. Zernik's approach brings together several current theories. RINA makes use of: theories of presupposition and semantics [Fillmore 81, Keenan 71], theories of dynamic memories [Schank 82], results on integrated parsing [Dyer 83], the knowledge based approach to natural language processing [Wilensky 81], unification grammar [Kay 79], and general learning strategies [Mitchell 82]. In RINA, Zernik also touches upon language generation problems as a validity test for the acquired knowledge. RINA learns through an interactive dialogue with a native speaker of English.

The program is composed of several basic components and their relationships are pictured in figure 3:

- The hierarchically organized phrasal lexicon: $DHPL$. It is the main data structure used in RINA and it is also the only learnable part of RINA's knowledge. $DHPL$ is discussed on page 7. There is also a simple word lexicon as an appendix to $DHPL$.
- A knowledge base of encyclopedic knowledge in the form of scripts, plans, and goals (SPG). This encodes world knowledge of many situations such as the organization of a trial, the role of school in human society, some historical/biblical facts etc. This knowledge base of memory chunks is neither learned nor modified by the program.
- A rule base allowing RINA to make inferences about the semantic situation an utterance refers to. These rules are numerous and are used in the process of acquiring a new English phrase or sentence. The rule database is not modified by RINA.

Figure 3: Flowchart of RINA

- There are also two components that operate on the preceding knowledge: the integrated parser using *DHPL*, and the single sentence language generator.

The program acquires lexical entries from context through a classical generate-and-test process. For RINA as for FOUL-UP, the context is the semantic situation referred to in the given utterance. It is first extracted, then the semantic role of the unknown lexeme is deduced a la FOUL-UP. Feedback is given by the user.

```
User: Al Capone went on trial, the judge threw the book at him.
RINA: He threw a book at him?
User: No, the judge threw the book at him.
RINA: He punished him severely?
```

Figure 4: A sample session of RINA

Figure 4 shows a sample session of RINA. In this example, RINA already has some knowledge about words used in the second part of the first user's inference, *"the judge threw the book at him"*. The entire idiomatic phrase is not in her lexicon and RINA first attempts to interpret it simply as a non idiomatic phrase, *"he threw the book at him?"* but the native speaker (*User*) rejects this interpretation. Since RINA has failed to interpret the sentence with the existing *DHPL*, she then tries to deduce it from the context. The context is in this case, the context of TRIAL-EVENT. It is represented by a script that contains information on the way trials are organized. RINA knows the role of a judge and is also aware of the fact that people like Al Capone should be kept in jail. This judgemental knowledge is of course pre-encoded by Zernik for the purpose of these example sentences. The main source of underlying knowledge is the TRIAL script that can roughly be described by the structure in figure 5 giving the basic events that take place with their relative order. From this context, RINA interprets the sentence as meaning, TO PUNISH HIM SEVERELY. Figure 6 gives the *DHPL* representation of the phrase *throw the book at.* We give more details on RINA's learning method in section three.

RINA's knowledge consist of two basic parts: linguistic knowledge on one hand, and world knowledge on the other. The two sources interact in the process of parsing, generating and acquiring English phrases and sentences. We review here these two sources of knowledge separately.

The linguistic knowledge: *DHPL*.

The basic structure used in RINA for representing linguistic knowledge is the Dynamic Hierarchical Phrasal Lexicon (*DHPL*). Entries are entire phrases hierarchically organized according to a relative measure of generality. RINA's main goal is the acquisition of English idioms. A sentence (or phrase) is represented as a more or less "frozen" structure. Parsing consists in unifying a given English sentence with *DHPL* entries, finding the most related structure. Idioms are

```
(a)     The prosecutor communicates MTRANS his arguments.
(b)     The defendant communicates his arguments.
(c)     The judge decides SELECT-PLAN either:
    (1) Punish (thwart a goal of) Defendant.
    (2) Do not punish him.
```

Figure 5: Sketch of the trial-script of RINA, from [Zernik 87]

```
PATTERN          ?x:person  throw:verb  <the book> <at ?y:person>

PRESUPPOSITION   ($trial (prosecution ?x) (defendant ?y))

CONCEPT          (act (select-plan
                        (actor   prosecution)
                        (plan (ulterior-crime
                                  (crime ?c)
                                  (crime-of ?y)))))
                 (result  (thwart-goal
                             (goal ?g)
                             (goal-of  ?y)))
```

Figure 6: Representation of *throw the book at*, from [Zernik 87]

traditionally recursively defined as "... *an expression whose meaning cannot be accounted for as a compositional function of the meanings its parts have when they are not part of idioms*" [Cruse 86]. Zernik breaks the circularity in claiming "equal rights" to idioms as part of a hierarchical phrasal lexicon. Idioms, dead metaphors, collocations and *regular* phrases are hierarchically organized in the lexicon according to their semantic opacity. The lower a phrase in the hierarchy, the more frozen its syntactic structure. Figures 6 and 7 are example entries in *DHPL.* The representation in figure 7 for the phrase *X command Y to Z* says that X presents an authority to Y, X tells to Y that Z is a goal of X. RINA is also given a word lexicon as an appendix of *DHPL*, where words are entered along with their syntactic category and part of their semantic traits. The choice of the semantic primitives, and their number is dependent on the knowledge of the world assumed by the program. It is largely arbitrary as can be seen in figure 8, which is the dictionary entry for *jogging.*

```
(comment    X command Y to Z)
(pattern  ((subject ((instance ?X)))
           (verb  ((root command)))
           (object1  ((instance ?Y)))
           (comp
             ((pattern ((subject ((instance ?Y)))
                        (verb  (form infinitive) (comp'er to)))
              (concept ?Z)))))
(presupposition (head authority)
                (high ?X)
                (low ?Y))
(concept  ((head MTRANS)
           (actor ?X)
           (yo ?Y)
           (obj ((active-goal ?Z)
                 (goal-of ?X)))))
```

Figure 7: A *DHPL* entry for phrases for *command* in RINA, from [Zernik 87]

```
     (..  (pattern 'JOGGING concept (ACTIVITY-THEME name
'running-for-fitness) .)
```

Figure 8: A lexical entry for *jogging* in RINA

From a psycholinguistics viewpoint, RINA tackles a topical subject. Second language learners often stumble over idioms in their everyday exposure to $L2$. In this context, $L1$ rules are useless as the proper usage of idioms or dead metaphors only comes with practice. Treating idioms as "equal citizens" supports the need for a phrasal lexicon. In addition, using a phrasal lexicon allows handling the semantic constituency of language and at the same time the integration of syntax and semantics. However some problems remain.

It has not been demonstrated that a phrasal lexicon is sufficient to account for possible compositional relations in natural language. Building a lexicon with only phrasemes doesn't seem very natural in order to account for lexical knowledge. For example the relations between *thick* and *thin* or between *dog* and *bark* etc. are not accounted for in RINA. These words are separate entries with no indication on their possible affinities. Another objection to the use of phrasemes only is that if we exclude the case of semantically opaque phrases, such as idioms, phrases are neither semantic nor syntactic primitive components of a language, and as opposed to words, are not in finite number.[3] Having phrasemes may seem useful, but why stop at a this level, and not include discourse strategies and theories of speech-acts in a sentence-lexicon, text-lexicon, story-lexicon etc?

The encyclopedic knowledge.
Besides the script based memory which is very similar to that used in FOUL-UP, RINA makes use of plans and goals allowing her to reason about scripts. The knowledge base consists of production rules which reason on SPG. Its complexity notwithstanding, RINA represents semantic knowledge in the same way as FOUL-UP or CHILD.

In summary, RINA heavily relies on encyclopedic knowledge, little attention is paid to formal syntax and the knowledge acquired is more related to the encyclopedia than to the dictionary. RINA uses as basic representational tools, a phrasal lexicon and a world-events memory. Using this method, RINA cannot account for phrasemes that cannot be explained by a simple and single world situation such as example: *to take an exam, to take a break, to let out a cry, to throw up* etc. These phrases, though not semantically more transparent than idioms, represent major difficulties for second language learners and could be accounted for at the lexical level, as we will see later on.

As in [Zernik 87], RINA's lexicon has around 200 phrases. The encyclopedic knowledge of RINA consists of 5 goal situations, 3 scripts 4 interpersonal relations and 200 specific planning rules.

2.1.4 Semantic Knowledge as Case Frames

Although RINA, CHILD and FOUL-UP are very different, since they all use conceptual dependency, they have the same approach as SAM in their representation of semantic knowledge,

[3]This is only true in theory. In practice, it is reasonable to assume a maximum number of words by sentence and thus to also assume a finite number of acceptable sentences. However, this number is exponentially related to the number of words in the dictionary.

and in a more general way, they share the same model of language competence. This modeling method was initiated by Schank in [Schank 72]. In addition, RINA makes use of abstract goals and plans and CHILD uses a set of rules in order to model the behavior of a child. In their work, meaning is the main concern and a sentence is fully described by a conceptual structure. Syntax is restricted to the set of procedures that help in the mapping between words and world, and lexical knowledge is not directly accounted for. Scripts describe the world as series of events taking place in a particular situation. In addition, plans and goals take human psychology and reasoning abilities into consideration. Scripts, plans and goals (*SPG*) are representational tools for the encyclopedia. They represent world knowledge to be used for natural language processing purposes.

The three programs have a module equivalent to the APPLY module in FOUL-UP. In the process of parsing a sentence this module has the task of retrieving the associated script (plan or goal) to account for the adequate situation. Consider the following sentences taken from [Granger 77] and [Zernik 87]:

(1) *"A car swerved off route 69. The flyvver struck a tree"*
(2) *"John ran over a pedestrian. He failed to* `explain it away` *in court, and he went to jail."*

For sentence one, the selected script is the "car-accident-script", and in the case of sentence two it is the "trial script" pictured in figure 5. The script selection mechanism is a problem that is not really addressed in the three programs above. In sentence two , the triggered script is the "trial-script" because of the word *court*, but there is no reason why *jail* does not trigger a "jail script" or *pedestrian* does not trigger a "street-event script". This information and choices are coded by hand before the program actually runs.

In addition, by manipulating such "canned" pieces of knowledge with production rules, the programs are able to imitate understanding and learning of expressions pertaining to several different domains such as: *to bury the hatchet, to kick the bucket* etc. However they also inherit the classic problems of such techniques. Learnable knowledge is constrained by the number and complexity of the scripts used; the three systems cannot deal with unpredicted events since the semantic knowledge is not upgraded along the way. This makes learning rigid and brittle.

In the three programs making use of *SPG*, there is a direct mapping between sentences and *SPG* memory, Zernik in [Zernik 87] claims that *"A relatively small number of structures can represent phrases whose instances can be used across many domains"*. The problem is that in none of the programs described here, is this part of memory upgraded. The possibly acquired knowledge is thus limited by pre-conception. Before learning what an *elm* is, one must first hand-code a description of how a car accident happens and before learning what *to throw the book at someone* means, one must first handcode information on trials. *SPG* represent world events and their structure doesn't reflect language use. Rather, they introduce an intermediary supplementary level between real world and language.

Though characteristic of *SPG* representation based systems, these problems are not limited to them. Modeling the semantic situation is a very hard task, and extending the complexity of the structures used may not be the solution. Elhadad in [Elhadad 87] demonstrates the need to take linguistic, semantic and pragmatic situations together into account in order to properly tackle the problem of situation. The only kind of situation taken care of in RINA and FOUL-UP, is the semantic extra-linguistic situation, and this limitation forces the authors to hand-code knowledge on domains such as human relationships or human institutions (school, justice etc.) in order to obtain "meaningful" interpretations of English utterances.

In general, the main flaw in these three programs is that they are trying to acquire a knowledge without specifying a performance task that would validate the results. This problem is due to the nature of the knowledge they attempt to learn. Encyclopedic knowledge has not yet been formalized by linguists, which doesn't provide a good framework to work in. In contrast, syntactic knowledge has been the most investigated part of natural language. Formalisms have been proposed and grammars have been written in these formalisms.

2.2 Syntactic knowledge: Berwick

In contrast to Granger, Selfridge and Zernik, Berwick in [Berwick 85] describes a program based upon a totally different approach. Language competence is restricted to syntax and the goal of the program is to learn to parse English sentences in a finite amount of time. We first describe the program and the assumptions it is based upon, and we then discuss the use of production rules to encode linguistic knowledge.

2.2.1 The program

With the exception of Debili's program, see section 2.3, the other programs described in this paper do not directly account for syntactic knowledge. Syntactic knowledge is not explicit but rather duplicated and dependent on the lexical entry considered. In contrast, Berwick puts the emphasis on the syntax, and on the formal rules of English grammar represented as production rules.

His program's task is to acquire syntactic knowledge in the form of grammar rules. Based upon a theory of parameterized universal grammar [Chomsky 86] and other theoretical results, Berwick's program is able to learn formal rules of English grammar. His program, given simple examples of language utterances, outputs grammar rules needed by a deterministic parser of English based on PARSIFAL [Marcus 80].

The basic assumption is that language faculty or competence is dependent on what is called a language acquisition biological device (*LAD*) [Chomsky 75]. This supposedly innate *LAD* has the task of learning/acquiring a language when exposed to examples of its use. It is to be viewed as a biological inference engine that accepts as input English sentences. and that after a length of time, outputs a deterministic parser of English. The other part of the innate linguistic competence or biological linguistic ability is described by a universal grammar or a set of universals. Universal grammar is a kind of generalization or abstraction of all the possible natural language grammars. Each particular or language specific rule is accounted for by a parameterized universal. By properly instantiating the parameters, we obtain the grammar of a given language. This is precisely the role of the language acquisition device. According to this theory, this is the way human beings, and particularly children process linguistic information in order to acquire fluency in their native tongue. The language acquisition device and the universal grammar together form innate biological linguistic competence. We discuss them in more detail in section four on language acquisition.

Berwick's program is intended as a model of $L1$ acquisition by children, and its goal is to generate a deterministic parser of English. Actually, in only generating the grammar rules needed by the parser, it also adds to the innate or presupposed knowledge an inference engine (the parser) and a set of predicates with which to write the grammar rules. Berwick's work is very rigorous and gives good results within its framework. His work is a partial implementation of the theory of universal grammar, in the sense that it should give satisfying results in any natural language whether it be English or Diola.

There are four basic structures in Berwick's program: the parser, the lexicon, the rule base and the learning algorithm. They are organized as shown in figure 9.

Figure 9: Flowchart of Berwick's program

- **The parser.** Its engine is based upon Marcus's parser, and is similar to a regular LR[3] parser, with a control stack. The parser is an interpreter of production rules that stand for formal grammar rules. A typical parse consists of making a single left-to-right pass through a given input sentence, executing along the way all the triggerable rules. The output of the parser is an annotated parse tree. The parser has 4 primitive actions which are part of the rule left-hand-side, **attach, switch, insert, trace.** They operate on the data structures of the parser (the three cells and the stack) and they eventually build the parse tree of the given sentence. Each action takes into account specific syntactic features, such as: passive sentences by **insert**, imperative by **insert**, subject complement inversion by **switch** etc. Berwick has drastically customized Marcus's parser for his purpose. The result is simpler and more amenable to learning.

- **The rule base.** Though specifically tailored to the parser used, it represents a coherent subset of English grammar. The language in which it is written consists of the four basic actions and the elements manipulated are the data structures of the parser (three cells, and a stack) and the English words themselves, see figure 10. This results in an *a priori* limitation on the space of possible rules, since the rule language is itself very small, the initial set of all possible rules is accordingly constrained. After a successful full run of the program, the rule base consists of 100 rules, all acquired. These 100 rules represent the final knowledge of the program. Berwick says that on the average, 70 to 100 rules are acquired. Examples of parsing rules are given in figure 10.

- **The lexicon.** Though being of remote use, the lexicon encodes some semantic knowledge and is mainly used in the syntactic disambiguation process. Words are entered along with basic syntactic features (transitivity, gender, number etc.) and also a few semantic traits (such as animal, human etc.). Verbs are also given the semantic roles of their actants, see figure 11. The nature and use of the lexicon is obviously not of primary importance in Berwick's work and few justifications are given. Part of these lexical entries can be completed during the acquisition phase of the program. Words are also given a *CLASS* field, which seems to point to synonyms. No justification is given on the choices of the extra-linguistic semantic attributes. Semantic information is used for word selection and simple disambiguation. Berwick gives few details on the assumed role of semantic and pragmatic information in his program. He considers semantic and pragmatic knowledge to be equivalent. The focus is on the syntactic realization of thematic roles (subject = actor...). This partial account of encyclopedic knowledge is probably a weak point despite the aim of modeling syntax.

- **The learning component.** When the parser fails to trigger any of the existing rules, it is a comprehension failure. Since the program is only given positive input sentences (grammatically correct and simple sentences), the rule base must be augmented in order to correctly parse the sentence. This is where learning actually takes place.

Berwick's program has the twofold goal to acquire linguistic (syntactic) knowledge and offer a psycholinguistic account of First Language Acquisition. Though Berwick in [Berwick 86] is currently extending his model to Chinese and German, the fact that the program has not yet been tested on other natural languages, is a drawback.

```
   Item examined                  Feature to match against
current active node:              Major sentence
    1st buffer cell:              Auxiliary verb
    2nd buffer cell:              Noun Phrase
    3rd buffer cell:              none

Rule action:    SWITCH first and second buffer items.

         (a) English description of grammar rule.
```

```
Pattern: Current active node             Buffer
                                    1st     2nd    3rd
           [**c; * is S Major]    [=Aux]  [=NP]   [ ]

Action:   SWITCH

        (b) Abbreviated form
```

Figure 10: An example parsing rule, from [Berwick 85]

```
ARGUMENTS:              NP, +ANIMATE, S
THEMATIC ROLES:         Subject is AGENT;
                        Object is AFFECTED OBJECT;
                        Propositional complement is THEME
CLASS:                  convince, tell ...
```

Figure 11: Lexical entry for *persuade*, from [Berwick 85]

2.2.2 Syntactic knowledge as production rules

The entire problem of language acquisition is much too large to be simulated at once. As we have seen, relying on encyclopedic knowledge in order to figure out the meaning of lexemes (words or phrases) was one way to restrict the problem. This restriction however introduced some difficulties. Similarly, considering only the acquisition of syntax is a way to restrict the problem. In contrast to semantics, syntax offers the advantage of being well formalized and suited to computer application. Most grammar formalisms (transformational grammars, lexical functional grammars [Bresnan 82], generalized phrased structure grammars [Gazdar 82].) give a formal analysis of English sentences and are thus well suited to computer applications. This relative success of syntactic models in computer applications makes it a kind of knowledge that is easily tested and directly usable. Neither advantage holds for encyclopedic knowledge.

2.3 Lexical Knowledge: Debili

2.3.1 The Program

Debili in [Debili 82] describes a program whose aim is the disambiguation of French sentences. Though disambiguation is not directly related to language acquisition, we have chosen to present this program here because the techniques used are totally relevant to our purpose. For solving ambiguities, Debili acquires information from previously read texts. Among the many different kinds of ambiguities possible in natural language, Debili concentrates on two. The two kinds of ambiguities[4] handled are:

- **Lexical** ambiguities of the type:
  (1) *The lease* expires *at the beginning of the month.*
  Three different meanings of the word *expire* are possible: *dies, exhales* and *terminates.* But only one is correct in this sentence.

- **Structural** ambiguities of the type:
  (2) *Blond school children.*
  Here, the ambiguity lies in the fact that we don't know if *blond* is to be considered as a modifier of *school* or of *children.*

The most commonly used approach to both kinds of ambiguities in natural language processing systems (such as Berwick's, ELI ....) consists of introducing semantic traits or features. In sentence (1) for example, *lease* would be given the $< non - animate >$ feature, thus helping in the selection of the *termination* meaning of *expire*. In sentence (2), *children* would be associated the trait $< has - hair >$, thus selecting *blond* as a modifier of children rather than *school.* In both cases, it boils down to introducing semantic traits and checking the compatibility of traits among

[4]For both types of ambiguities, we use the terms chosen by Debili in [Debili 82]

the syntactic components of the utterance. This method has been widely used in natural language processing systems, and, though yielding interesting and attractive results, it suffers from its lack of generality in terms of the semantic decomposition. Conceptual dependency as used in RINA, FOUL-UP and CHILD is a direct application of this method. Many questions are still open such as: Where to stop in the semantic decomposition of the words? What to chose as primitive traits? Including $< has - hair >$ rather than $< doesn't - have - a - driving - license >$ or $< innocent >$ seems rather arbitrary and ad-hoc.

Debili takes a different approach. Instead of making use of referential or encyclopedic knowledge, he claims that lexical knowledge can be used efficiently in the resolution of the above ambiguities. As pointed out by Katz [Katz 80], there is more to language than a grammar generating well formed sentences and a semantic analyzer mapping sentences to real world propositions. A lot of links are missing in this schematic approach for the mapping between texts and meaning. One of these missing links is the set of rules describing the way words collocate in natural language sentences. Debili's goal is to use this knowledge in order to solve two kinds of ambiguities in French texts. This knowledge is also termed cooccurrence knowledge [Ducrot 79].

His program has a twofold goal:

1. Acquire a large number of collocations or dependency relations from natural language texts.
2. Use them in order to disambiguate sentences of type (1) and (2).

The program has three components: the morphological analyzer, the syntactic analyzer, and the semantic component.

The morphological analyzer works in two passes. First, from a given sentence $S = (w_1, w_2, \ldots, w_n)$ it derives the morphological sentence: $M = (e_1, \ldots, e_n)$, where e_i is equal to (w_i, s_i) and w_i is the i-th word in S, and s_i represents its syntactic attributes. Then, for each e_i of M, it derives the morphological stem of the corresponding word (ex: ambiguousness → ambiguous) and also the set of potentially related groups of words (ex: expire → death, exhale, terminate). The morphological analyzer uses a precise knowledge of French morphology compiled in a sort of thesaurus before the program actually runs.

The syntactic analyzer works on the output of the morphological analyzer, parses and at the same time identifies all the ambiguous and non-ambiguous dependency relationships between words of the sentence. For example, in sentence (2), the relation *school-children* will be termed unambiguous, whereas both *blond-children* and *blond-school* will be declared type (2) ambiguous.

The semantic analyzer works in two stages. First it stores all the relations declared unambiguous by the parser. This is where the actual acquisition takes place. Then it tries to solve the ambiguous ones by using the permanent memory of previously stored dependency relations.[5] Only ambiguities of the types (1) and (2) can efficiently be solved. Debili says that in a typical technical French text, on the average of 80% of the dependency relations are unambiguous.

[5]The order of formation is not taken care of.

and that after the semantic disambiguation, one third of the ambiguities are solved. No indication is given as for the remaining ambiguities.

2.3.2 Lexical Knowledge

The fact that *lease* and *expires* appear together (or cooccur) in utterance (1) is obviously not only due to the fact that "the *lease expires* .." is grammatical and that *lease* and *expires* have compatible semantic features. They share what is called a lexical affinity [Halliday 76]. Debili has defined and used eight types of such affinities appearing in French sentences. He calls them: Lexical Semantic Relationships or *RLS*[6].

These relationships (or affinities) are of two kinds: paradigmatic and syntagmatic [Saussure 49]. There is a syntagmatic affinity between two units of language w_1 and w_2 if there is a correlation between the appearance of w_1 and w_2 in the utterances of the language. Similarly, there is a paradigmatic affinity if w_1 and w_2 can replace each other in the same utterance without distorting the meaning too much. Paradigmatic and syntagmatic affinities are also termed metaphoric and metonymic affinities [Ducrot 79]. To give an example, *cat* and *feline* share a paradigmatic affinity whereas *cat* and *graceful* share a syntagmatic affinity.

Debili partially accounts for both kinds of affinities in his program. Paradigmatic affinities are stored in the thesaurus (ambiguous, ambiguity, ambiguousness, etc..) and syntagmatic affinities are acquired and stored in the *RLS* memory. Eight types of syntagmatic *RLS* have been identified by Debili. Their classification is done according to the part-of-speech of the two units involved. For example there is an *RLS* of type 1 between *boy* and *give* ("the boy gives"), and between *dog* and *run* ("the dog runs"). Similarly, there is an *RLS* of type 2 between *big* and *cat* and between *blond* and *children* ("the big cat" and "blond children"). Type 1 stands for noun-verb relations, and type 2 for adjective-noun.

A lot of research has been done in linguistics towards the formalization of these affinities [Halliday 66], [Mel'čuk 81]. Debili's account of lexical cooccurrence is restricted and suffers from a weak formalism, but demonstrates one of the possible use of lexical knowledge in computer programs: surface disambiguation. In Debili's work, most of the information gathered in the form of *RLS*s is not relevant to a linguistic description. An example will clarify our point. Suppose that the program parses sentences such as:

1) *The poor fox had a narrow escape yesterday*
2) *This fox is rather narrow*

The program will store the *RLS* *narrow-escape* and *fox-narrow* as a first stage. The point we would like to make is that although a fox can be narrow, the link between *narrow* and *escape* is a real syntagmatic affinity, whereas *narrow* and *fox* just have compatible features.

The program is unable to distinguish between linguistically useful and irrelevant information.

[6] From the French word ordering.

One problem here with lexical knowledge is that it has not been compiled in any sort of dictionary or grammar book for English [Markowitz 86], so that it is hard to check the validity of the acquired knowledge. In chapter five, we describe a linguistic theory making explicit use of lexical knowledge that gives a precise and extensive description of lexical cohesion, allowing filtering of irrelevant data.

3 The Learning Methods: a Comparison

As mentioned earlier, learning ability, learnable knowledge, as well as the learning methods and results are mainly determined by the knowledge representation techniques chosen. Therefore it comes as no surprise that the systems are partitioned into the same groups in this chapter on learning as in the previous one on knowledge representation.

3.1 Expectation Based Learning

The basic strategy of expectation-based learning, the method used by the programs acquiring encyclopedic knowledge, can be described as the process of filling up expectations based upon the semantic situation referred to. This method is inspired from the expectation-based parsing method, already used in *ELI* [Riesbeck 76]. The encyclopedic context triggered by a key word is the basis of the learning process, it uses pre-encoded semantic knowledge to guess the meaning of the unknown word. Expectation based learning in the context of language acquisition is best represented by RINA (FOUL-UP and CHILD are also based upon the same mechanisms). When in the presence of an unknown word, the world memory (described by *SPG*) is used to derive semantic information from the sentence. Then the role/meaning of the unknown word is guessed from the memory expectations. This method attempts to find a mapping between a lexeme (words for FOUL-UP and phrases for RINA) and a particular world event or situation. Computer implementations of encyclopedic knowledge are less successful than those using syntax only. When compared to syntactic knowledge, encyclopedic knowledge is less formalized. Learning it is harder. and testing the validity of the acquired knowledge is difficult. Moreover, the learning task is less easily describable as a general learning task. To our knowledge, work in the machine learning field has rarely been applied to the acquisition of encyclopedic knowledge.

In the rest of this section, we focus on RINA although the same could be said for CHILD or FOUL-UP. Descriptions of RINA are more detailed than the other programs and this makes things clearer and easier.

RINA uses feedback from a native speaker of English, thus being provided with a type of near-miss. When RINA's guess has been denied by the user as in figure 4, then RINA tries to explain the phrase using context or already existing concepts. Assume that RINA is provided with the following examples taken from [Zernik 87]:

David took on *Goliath*

The Celtics **took on** *the Lakers.*

In this case RINA knows what happened between the Celtics and the Lakers, and between David and Goliath. In both cases, the subject of the sentence has beaten the other in a rather courageous manner.[7] RINA deduces the meaning of *take on* as being the common relation between the Celtics and the Lakers and between David and Goliath.

This strategy, though representative of part of human processes, is not fully satisfactory. The amount of knowledge to be pre-encoded would rapidly get out of hand for a general purpose computer application. Moreover, as we said before, though encyclopedic knowledge must take part in a full-fledged model, it has not yet been successfully applied in a general purpose computer application. This makes it even harder to learn for a computer program. Progress should first be made on its formalization.

3.2 Rule Acquisition

The problem of acquiring English grammar has extensively been investigated in the three fields of computer science, psycholinguistics and linguistics. Learnability of natural language in this framework can easily be expressed in terms of search space reducibility. However, using a general formulation, in which innate knowledge consists of the *LAD* and the universal grammar, some grammars cannot be learned in a finite amount of time [Grimshaw 84] and thus the search space must be reduced. In addition to Berwick in [Berwick 85], this problem of acquiring an adult grammar has also been tackled by Vanlehn in [Vanlehn 87] and Berwick in [Berwick 87]. In these works as well as in Berwick in [Berwick 85], the reduction of the problem is classically done in one of two different ways: either put a lot of knowledge in the *LAD*, or constrain the search space of the grammar rules in some way. In both cases as pointed out by Langley in [Langley 87]: "*... motivation for studying grammatical inference is completely independent of psychological and linguistic issues - it holds significant interest as an abstract learning task*". I would add that since grammar is already formalized and compiled in grammar books in a way that has already been used by computer programs (*e.g* [Marcus 80]), the performance task is not the main motivation either.

Berwick's program is based on two learning strategies that greatly reduce the search space. The first one, the **subset principle** is a general learning technique. The second one, $\bar{\mathbf{X}}$ theory is a linguistic assumption. We review both of them in turn, and we then examine the achievements of the program.

The subset principle is a refinement of the version-space learning algorithm [Mitchell 78] in the context of learning from example with positive-only examples. The search space is the set of all possible grammar rules. The subset principle applies the general heuristic that considers that the set of plausible hypotheses (or the version space) is the most specific set of concept descriptions that is consistent with the training instances seen so far. In the case of positive only examples, Berwick in [Berwick 86] calls the subset principle a general learning theory for

[7] Actually, since then, the Lakers have taken revenge.

language, and says that "*it drives language through its stages ... and it is sufficient to guarantee successful acquisition after some finite number of positive examples.*" The fact that the program works with positive-only examples, is part of the psychological validity criteria of the model described by Berwick.

The second theory which is part of the foundation of Berwick's program is $\bar{X}$ theory. In stating that there is always a fixed ordering of sentence substructures in most natural languages, $\bar{X}$ theory drastically reduces the set of possible context-free rules of a language. For example there is no verb phrase starting with a preposition. Phrases are built around a central scaffolding that includes a basic *head* item, a *specifier* and a *complement.* Learning language structures is thus made easier, since for example, languages are either head-first (like English) or head-last (like German). By constraining the search space in such a way, Berwick has been able to acquire a large set of rules in the form of parsing production rules.

Berwick's first claim is that, given a set of universals, the subset principle, $\bar{X}$ theory, and simple input sentences, his program learns a large subset of the formal grammar rules of a natural language. However, the actual relation between a given set of universals and the set of rules defining a language is not clear [Rutherford 84]. The mapping between the theory and the implementation is not clear either in the descriptions of the program. For example the claim that the search as implemented is language independent (at least before the program is run) is not convincing as long as the program has not been tested on several other languages.

Berwick's second claim is that his program is an account of $L1$ acquisition by children. But formal grammar represented by a set of rules, does not account for all parts of language and cannot be considered an $L1$ acquisition model. Acquisition of grammar by children is deeply entangled with many other aspects of education such as the acquisition of semantic and pragmatic knowledge, or the individual's cognitive and social development. In summary, the program cannot be considered a full model of $L1$ acquisition but rather as a partial implementation of theoretical assumptions about universals.

3.3 Systematic Acquisition of Lexical-Semantic Relationships

Debili sets simpler goals for his program: doing rote-learning of lexical-semantic relations from context. In his work, each non-ambiguous syntagmatic affinity (between any pair of words) that is identified at the parsing stage, is definitely entered in the RLS-memory. A large amount of information is stored, and a great part of it is irrelevant to a description of language activity. RLS between *wind* with *blow* or *dies down*, as in "*the wind blows and it dies down*" are necessary to the proper usage of the word *wind*, whereas others are not worth storing. Debili offers no mechanism to select RLS entries according to their linguistic interest, thus keeping in memory a huge amount of insignificant RLS.

In summary, Debili's program is to be considered as a first attempt to apply language acquisition to cooccurrence and use it for disambiguation. Other attempts to use cooccurrence knowledge have been made in language generation and paraphrasing [Boyer 85], and computational lexicography [Markowitz 86, Smadja 88]. Progress can be made in systematizing this information and in drawing from theoretical linguistics on cooccurrence knowledge.

3.4 Synthesis

In summary, language acquisition is a hard subject that cannot be tackled in isolation from linguistic and psycholinguistic aspects and that cannot be considered as a single indivisible process. All five programs described above have attempted to identify subtasks and to isolate subproblems. For the sake of clarity, we re-examine the five programs here, in light of linguistic constraints.

From a **linguistic** viewpoint, we have distinguished among four kinds of knowledge used:

- *Syntactic knowledge* or knowledge of formal grammar. This knowledge has the advantage of being well formalizable and suitable for a computer model. It has been the focus of attention in linguistics for many years. Since it has a well defined and testable performance task, it is a good learning task, although the acquired knowledge is often not the motivation in itself. This is the approach taken by Berwick in his program. Berwick gives indications on how to evaluate his program and also gives an evaluation of his results.

- *Encyclopedic knowledge* or extra linguistic semantic knowledge. Here also a lot of work has been done ([Jackendoff 76], [Fillmore 78],...) but very few models have been formalized precisely enough to be successfully used by computer systems. See Section 4.2. This is a major drawback for learning this knowledge, since without a performance element, a learning program cannot be verified or evaluated. None of the three programs are evaluated by the authors.

- *Lexical knowledge.* This type of knowledge has only recently been used in natural language processing systems and several models have been set forth by linguists. However, it is generally overlooked not considered in current works. Debili attempts to acquire lexical knowledge in the form of *RLS*. Some arbitrary approximations are made, and a lot of insignificant knowledge is stored. Debili gives precise figures and evaluation tests of his program and although the task is restricted, the results seem satisfactory according to the performance task (surface disambiguation).

- Other kinds of knowledge, such as *situational, pragmatic, psychological, etc.* CHILD is the only system to make explicit use of psychological knowledge. The other kinds of knowledge have not been investigated though some formal models are available ([Allen 80, Searle 69, Barwise 83] ...). Psychological knowledge is the least formalized, and as such forces Selfridge to make simplifications that are drastic and arbitrary to our opinion.

Given what has already been done, one direction of research would be to adapt and apply a semantic theory (or a pragmatic theory) adequately and apply it to learning. However it seems much safer and more natural to start by searching in the direction of lexical theories, since this is easier to handle and test than encyclopedic knowledge. In chapter five, we describe linguistic work on lexical knowledge and show how it could be used in the framework of language acquisition.

From a psycholinguistic viewpoint, if we aim at making a cognitive model of language acquisition, then we must either account for the full range of language activities or identify a subset of language competence and show that it corresponds to a psychological reality for language learners. For example, designing a program that learns the meaning of words starting with the letter *"a"* is not satisfactory for a cognitive model. Whether syntax is a psychological reality is strongly debated but in any case, first language acquisition cannot be separated from its social and psychological/cognitive environment. It can hardly be simulated without drastic and biased simplifications. Among the five programs described above, four (all except Debili's) claim to account for cognitive processes, and RINA is the only one dealing with second language acquisition. An overall classification of the programs described in this paper is given in figure 16. In the upcoming section, we review research work on the field of language acquisition and distinguish between first and second language acquisition.

4 Language Acquisition

The problem of language acquisition has been studied in various theoretical and applied fields such as linguistics, applied linguistics, pedagogy, psycholinguistics, lexicography and also computer science. In all these domains, the loci of interests are different and the terminologies employed often overlapping. This has resulted in apparent chaos, where a proliferation of similar terms have co-existed taking on different connotations depending on the theoretical standpoint.

For example, talking of *acquisition* rather than *learning* is more than a stylistic matter, but is rather a way to render a certain commitment of the writer. In computer science, talking of *language learning* is a way to refer to the field of Machine Learning, where the learning of a language is considered as the application of general learning strategies to linguistic input. Berwick and Vanlehn used this approach. In contrast, Zernik uses the term *language acquisition* and does not make use of general machine learning techniques. In lexicography and applied linguistics, the preferred term is *learning* [Cowie 78], whereas in psycholinguistics it is *acquisition* [Lyons 68], [Krashen 82], [Hyams 86]. Krashen, a psycholinguist, defines his own interpretation of the twin concepts and builds a model of second language *acquisition* around this distinction. In linguistics, where the usual term is *acquisition*, Halliday in [Halliday 75] takes the other tack in choosing *learning*; talking about language *acquisition*, he says, is a way to think of a language as an external skill or property to be acquired as such. In this paper, our preferred use of *acquisition* should not be perceived as a commitment, but rather as the expression of non commitment, since *learning* has additional connotations in computer science.

In the previous chapters of this survey paper, we have described work done in computer science on the subject of language acquisition. This chapter is devoted to related work in the fields of linguistics and psycholinguistics. We discuss here the main trends. We first present work done in the framework of the Chomskyan linguistic theory. We then introduce the principal results of psycholinguistics in word meaning acquisition and second language acquisition (SLA) research.

4.1 The Chomskyan Viewpoint: to Grow a Language

Within the numerous subtasks and environmental conditions involved in the process of acquiring a language, the focus of attention in linguistics has often been restricted (whether implicitly or explicitly) to the acquisition of the syntactic component of the grammar of the mother tongue by children. The dominant standpoint is to consider the acquisition of linguistic structures as the heart of the acquisition process. Most of this work is done within the framework of Generative Grammar (GG) and government binding theories [Chomsky 81], where language is considered as a mental organ growing in the mind when fed with linguistic input. Children neither acquire nor learn the language but rather grow it. They are *a priori* provided with a Language Acquisition Device (LAD) and a parameterized universal grammar whose parameters are fixed along the way in order to produce an adult grammar. This approach is sometimes referred to as *the logical problem of language acquisition.* The development of children's linguistic system is compared to the development of their auditory or visual systems. In the same way that a child totally deprived of sound stimuli could never learn how to hear, if totally deprived of linguistic input s/he could never learn how to speak. This nativist approach has been widely discussed even within the framework of GG [Comrie 82], [Greenberg 78]. Wanner in [Wanner 82] gives the current directions of research for this theory and [Baker 82] extensively discusses its relevance to first language acquisition.

The logical problem of language acquisition has been studied in a theoretical way by Grimshaw in [Grimshaw 82]. She revised the model proposed by Chomsky and augmented it with an innate cognitive capacity aiding the LAD in its decisions. Hyams in [Hyams 86] elaborates on this basis in investigating evidences for an interactive model of language acquisition. She studies in detail the different developmental stages of children's grammars. She views grammatical development as a continuous succession of intermediate grammars, stretching from the inborn universal grammar, to the adult steady-state grammar.

Surprisingly, whereas FLA is hardly distinguishable from other processes (such as: cognitive, sociological, etc.) Halliday's linguistic research on FLA, [Halliday 75], has been directed towards the comprehension of the acquisition of syntax. This is probably a result of the fact that most of $L1$ acquisition research is performed in the framework of GG where the first and foremost motivation is syntax. In contrast, psycholinguistic research has focused on diverse parts of the acquisition process; we review in the next two sections research done in the fields of lexical development and SLA.

4.2 The Acquisition of Word Meaning

Whereas linguistic research has mainly been concerned with the acquisition of language competence by children, its psycholinguistic counterpart is more directed to the study of language performance. Because of the distinction between competence and performance made by Chomsky [Chomsky 57], psycholinguistic studies of language acquisition are often based on experimental data and result in models and strategies for pedagogy and lexicography.

Faced with the riddle of young children's lexical development, researchers in psycholinguistics and psychology have extensively investigated this subject. Their efforts have reached what can be termed a descriptive adequacy: the successive developmental stages children go through while acquiring a language have been spelled out and described abundantly in the literature. This work will not be discussed here, but the reader should refer to [Clark 77] for a survey on the state of the art of the subject. The second line of research has attempted to reach what can be called explanatory adequacy. Researchers have proposed several models accounting for various parts of the experimental data gathered by psychologists. The models attempt to explain how young children organize semantic information and what are the process(es) that allow them to constantly acquire new words. The main contributions of these directions of research are briefly reviewed in section 4.2.1, and, section 4.2.2 presents an evaluation of the work done in the area of word meaning acquisition from the viewpoint of search space and learnability.

4.2.1 Semantic Information and Mental Modeling

Children between two and five are able to learn a new word every waking hour [Carey 78]. Given this rate and the fact that encyclopedic and linguistic knowledge are deeply interrelated and constantly feed off one another, the question arises as to what kind of model is available to young children. Characterizing children's first words is a curious puzzle as children are at the same time augmenting their dictionary and encyclopedic knowledge. Previously acquired words are constantly modified and upgraded as new words or situations are encountered. Clark in [Clark 74], Nelson in [Nelson 79] and Rosch in [Rosch 75] have each proposed a model accounting for children's first words and explained how encyclopedic knowledge evolves along with the size of the lexicon. Their three models constitute the main coherent approaches to the problem. They first propose a model of lexical semantic information and describe how this semantic information is successively upgraded along the way. They disagree as to the nature and organization of semantic information, but account for basically the same phenomena: the interactions of lexical and semantic information in young children. The three models we discuss here are: the Semantic Feature Hypothesis (SFH) [Clark 74], the Functional Concept Hypothesis (FCH) [Nelson 79], and Rosch's Prototype Hypothesis (PH) [Rosch 73,75].

Clark's Semantic Feature Hypothesis

For Clark, each word is represented by a set of semantic features *à la* Katz in [Katz 72]. These semantic features comprise the meaning of a given word for a child. They are generally not identical with the set of features comprising the meaning of the same word in the adult language. Clark describes how this encyclopedic semantic information is dynamically modified as the child learns new words or concepts. She calls these successive modifications, *semantic restructuring.* Figure 13 gives an example of semantic knowledge restructuring. At the first stage, the child has already learned the word *bow-wow.* However s/he doesn't use it appropriately but rather generalizes its meaning to all four-legged animals. W_1, W_2 ... are the other words existing in the child's lexicon at that time, and the FS_i are their associated sets of semantic features[8]. Now let us assume that in the meantime the child has encountered a non-dog animal, such as a cow. At this point, the child realizes that s/he has been overgeneralizing the use of *bow-wow* and restructures his/her lexicon with the help of a new semantic feature, namely: *moo*. The lexicon then undergoes a restructuring and as can be seen in figure 13, *bow-wow* now includes the feature -*moo* whereas *moo* includes +*moo*. Clark's model accounts for such a dynamic lexicon in a rather precise way. The main criticisms of Clark's model may be on the choice of the semantic features used and the assumptions it implies[9]. No real methodological framework is given and the model is not clearly linked with psychological results on perceptual development [Atkinson 82]. However, Clark provides solid guidelines on which to base a computer model, however, to our knowledge, SFH has not been used in computer science.

```
Bow-wow     <->       + 4-legged
W1          <->       FS1
W2          <->       FS2

.....                 .....
```

Figure 12: Child lexicon at stage one. From [Clark 74]

```
Bow-wow     <->       + 4-legged, - moo
moo         <->       + 4-legged, + moo
W1          <->       FS1
W2          <->       FS2

.....                 .....
```

Figure 13: Child lexicon at stage two, after restructuring. From [Clark 74]

Nelson's Functional Concept Hypothesis

Nelson's approach is very similar to Clark's and as such, is subject to the same criticisms. The main difference is in the nature of the semantic features, which are of a more general nature in Nelson's work. Figures 14 and 15 show the semantic representation of the word *ball* allegedly used by a child, at two consecutive stages. The first stage represents what the child has encoded after his/her first experience with a ball. The second stage is what results after the restructuring

[8]This example is taken from Clark's work.

[9]For example assuming that a mental representation is based on semantic features. In contrast, Rosch's model is not based on this assumption.

(functional synthesis in Nelson's terms) of the concept after the child has had a new experience with a ball. Like SFH, FCH offers no real justification as to the nature, number and values of the semantic features. The other main problem is that restructuring the encyclopedic knowledge is extremely complex[10] and would take a very large amount of time for a large enough lexicon. As a computer model, FCH would be hard to use and leave the designer with a lot of simplifications to be done. The theory is far too unconstrained and unjustified, Nelson's model is based on Clark's one and it adds unjustified extensions to it. Atkinson in [Atkinson 82] terms FCH of "little interest for lexical development."

```
In living room, porch
Mother throws, picks up, holds
I throw, pick up, hold
Rolls, bounces
On floor, under couch
```

Figure 14: Meaning of *ball* at stage one. From [Nelson 79]

Location of activity:	Living room, porch playground
Actor:	Mother, I, boy
Action:	Throw, pick up, hold catch
Movement of object:	Roll, bounce
Location of object:	On floor under couch, under fence

Figure 15: Representation of *ball*, after the functional synthesis. From [Nelson 79].

Rosch's Prototype Hypothesis
As opposed to SFH and FCH, Rosch's model is not based on the assumption that words are represented as sets of features. Rather it draws on what is called the prototype hypothesis (PH) [Rosch 73, 75]. The PH states that semantic categories admit degrees of membership. It is based

[10] No indication on its space or time complexity has been found in the literature, and the algorithm is not described precisely enough for the author to deduce it.

upon a long series of experiments on the adult system of semantic categorization. According to these empirical data, it seems that semantic categories in adults are not defined by a set of criterion, but rather have members which are more or less "good members" or prototypes in Rosch's terminology. To give an example, *robin* is a prototype of the category *bird* whereas *penguin* is not. A prototype can be defined as a member of a semantic category that can easily substitute for the name of the category itself in most sentences. For example, replacing *bird* by *robin* in the sentence: *"I heard a bird twittering outside my window"* seems natural, whereas replacing it with *penguin* or *ostrich* is rather odd. A lot of research has been performed on the applications of *PH* to language acquisition issues [Heider 71], [Reich 76], [Bowerman 78] etc., but a well defined model of early lexical development is yet to be set forth. However, empirical data have been gathered [Griffiths 76], [Bowerman 78] that indicate the major role played by prototypes in lexical development of young children. Substantial discussion on this subject can be found in [Atkinson 82].

We conclude this section by a final critique on some of the computer programs described earlier in this paper. Among them, three were modeling word-meaning acquisition: RINA, FOUL-UP and CHILD. Of the three, none used any of the models presented above, the three prefer to rely on their own judgment and build and implement their own model. Zernik simply uses the fact that semantic concepts are hierarchically organized. And Selfridge justifies his departure from Rosch's *PH* in saying that his aim is not to model children language learning capabilities in general but rather to "... *describe the acquisition of particular concepts by a particular child...*" which appears to us as weak reason. We now turn our attention to the learning mechanisms involved in the acquisition of lexical items, and we consider the computer modeling of this activity.

4.2.2 Search Space, Learnability and Lexical Development

Acquiring a new word is an inductive task, and as such, it should be expressible in terms of search space and should be analyzable in light of the related work done in artificial intelligence. Unfortunately, none of the models discussed above offers a search space with a finite branching factor [Carey 83] and very little work has been done in this direction. An example will help clarify our purpose: Suppose that you are communicating with a non-English speaker and that at one point, the informant points to a rabbit and says *gagavai*[11]. You immediately hypothesize that *gagavai* means *rabbit* and you might want to test your hypothesis by pointing out other animals and rabbits and querying for each one: *gagavai?*. Suppose then that the informant demurs for each non-rabbit animal and ascents for each rabbit, you would probably feel more and more secure about your initial hypothesis and conclude that *gagavai* means *rabbit*. However, at this point, the number of remaining possible meanings for *gagavai* still remains infinite and can hardly be reduced by the responses of the informant. Indeed, *gagavai* could as well mean $< rabbit\ or\ ladder >$ or $< rabbit\ with\ non - green\ hair >$. Such terms are called "unnatural kind concepts" in opposition to "natural kind concepts" [Keil 81]. The rabbit example above, illustrate the fact that even though the only things pointed out by the informant are a rabbit and an elephant (by example), the concept associated to the word *"rabbit"* might be of any

[11] This example is drawn from [Quine 60]

form. Unnatural kind concepts are possibilities that are automatically ruled out by children or adults. In an explanatory theory of lexical development, or in a computer model, the search space must be reduced. Very few researchers have looked for natural constraints ruling out unnatural concepts [Atkinson 82], and the evidence so far seems to indicate that until the contrary is proved it can be assumed that there is none [Goodman 55, Carey 83]. In the framework of syntactic acquisition, such constraints have been proposed in the form of universal grammar, but in the domain of lexical development, this lack of constraints has forced Zernik, Granger and Selfridge to rely on their own judgment in making simplifications. This problem is a real impediment to computer modeling and greatly reduces the motivation for making a computer model of lexical development before more psychological evidence has been provided.

The next section presents research done in the field of second language acquisition where the general viewpoint comes from the domain of applied linguistics.

4.3 Second Language Acquisition (*SLA*)

Language acquisition research distinguishes between first language acquisition, second language acquisition and re-acquisition of a language (when a language formerly learned is first forgotten then re-acquired). Excluding cases of early bilingualism, it can be said that the first language is normally acquired during childhood whereas a second language is acquired when the individual has passed age 8. This means that the main difference between *SLA* and *FLA* is that *FLA* is deeply intertwined with the social and cognitive development of the child. Second language learners can focus on the acquisition of a more linguistic type of knowledge. In other terms, the task of the second language learner is more linguistically oriented that that of the first language learner. Thus, when studying language acquisition, it is easier to isolate linguistic phenomenon by studying the behavior of second language learners.

Although there have been theoretical investigations on what could be termed the logical problem of second language acquisition [White 83, Adjemian 84, Rutherford 86], the thrust of the work is more applied. Several models of *SLA* have been proposed. They are all based on common reflections on the subject. Among them, two basic points are still much debated amongst researchers and determine the orientation of the models proposed:

- The first is the distinction between the processes involved in *FLA* and *SLA*. The *identity hypothesis* asserts that the fact that a language $L1$ has already been learned or acquired is of little importance for the acquisition of $L2$ and the *contrastive hypothesis* asserts the opposite.
- The second is that several models of language acquisition have distinguished between explicit and implicit knowledge available to language learner [MacLaughling 83, Byalistok 78, Krashen 82]. Those two kinds of knowledge are roughly defined as the "knowing what" and the "knowing how" of language.

Rather than making a comprehensive survey here of all the current theories, we have chosen to discuss the above points in more detail. For an extensive survey on the young field of *SLA* the reader should look at [Klein 86], and for a survey of those theories from the viewpoint of the comprehension processes involved the reader should see [Walters 87].

4.3.1 Contrastive and Identity Hypotheses

A classic point in favor of the contrastive hypothesis is that communities with the same linguistic background seem to make the same kind of errors. In other terms, $L1$ seems to be directly influencing and modeling the learner's linguistic behavior in $L2$. Magiste in [Magiste 85] specifically studies the intra and interlingual interferences of several $L1$s and $L2$s.

Besides direct influence as studied by Magiste, the only fact that one already speaks a language greatly influences the process of language acquisition. For deriving information or understanding an utterance, one uses different sources of knowledge such as world, situational and pure linguistic knowledge. At the very least, we can say that a second language learner already possess the first two sources. Thus the task of the second language learner is considered to be more linguistically oriented than that of a first language learner. Even just considering the acquisition of linguistic knowledge, a second language learner is much more likely to master certain language features sooner than a child in the period of learning his/her mother tongue. For example, a second language learner is capable of understanding the basics of deixis[12] which are known to function nearly in the same way in most languages. The same goes for ellipsis, anaphora or other language components where the task of the second language learner is simply to understand the form and not the function.

Bailey in [Bailey 74], sets forth an hypothesis based on empirical results that should contradict the contrastive hypothesis, the *natural order hypothesis*. She shows that adults and children with different backgrounds, use common strategies and process linguistic data in fundamentally the same way. A language learner traverses several stages in his/her acquisition of language syntax, and the ordering of these stages has been showed to be (for second language learners) mostly $L1$ and age independent.

There is a common intuition that the identity hypothesis is true. In support of this, people often believe that children are more gifted than adults at learning languages. This belief has been seriously questioned since, a child is supposed to have fluency in his/her first language at age 3 but doesn't speak before age 1 or 2, which represents a huge amount of exposure to the language. Second language learners, however are required to output (although simple) sentences in $L2$ after a few weeks of learning (roughly less than 100 hours of exposure). That is, the manner and length of exposure could account for the difference between adults and children in their language acquisition skills.

There is no agreement about the identity hypothesis as stated above in its extreme form, but

[12]Elements of reference such as *today, tomorrow, here, there, this, that, now, then, before, etc.*, which serve to locate what is being referred to in space or time relative to the time or place of utterance, are known as deictic elements, and the phenomenon in general is known as deixis. [Cruse 86]

some authors do agree that there is a basic common chore in FLA and SLA. There are a lot of arguments in favor of both assumptions. As an example, every second language learner tries to apply $L1$ rules during the acquisition of $L2$. This phenomenon called *transfer* has extensively been investigated and reported in the literature. The claim was that the comparison of the structures of the languages involved would help predict the learner's encountered difficulties. However, no definitive conclusion has yet been drawn that could support one or the other hypotheses.

Neither of the two hypotheses can be rejected nor taken for granted by now. The difficulty is that there are both similarities and differences between FLA and SLA, and both forms should be studied independently and in connection in order to investigate the possibility of a single theory/model of language acquisition.

4.3.2 Explicit vs. Implicit Knowledge

In addition to the previous debate, a second direction of research has emerged in SLA research. Its main point has been to separate out the processes of guided and unguided language learning. Krashen's model [Krashen 82] takes root in the distinction between learning and acquisition for language learners. What Krashen means in opposing both terms is that there are two sorts of language knowledge available to the learner. The first kind of knowledge, the learned knowledge, contains the conscious facts that the learner knows about $L2$, such as some grammar rules, pronunciation rules, some words recently seen etc. This knowledge is also called explicit knowledge by Byalistok in her model [Byalistok 78]. The second kind of knowledge, the acquired knowledge, or implicit knowledge in Byalistok's terminology, contains the implicit intuitive and spontaneous information about the language. The acquired knowledge is generally acquired from participating in conversation.

Couched in somewhat different terms, Krashen's theory is based on the distinction between guided and unguided (spontaneous) SLA, the first controlling the latter. Byalistok [Byalistok 78] also describes a model of SLA based upon the same distinctions. However, her work is geared towards the description of the cognitive processes involved. Both models put the emphasis on the gap between the implicit and explicit knowledge available to language learner in general and to second language learners in particular. To acquire fluency in a language can be partly described as a transfer of knowledge from the explicit knowledge base to the implicit knowledge base, and both Krashen and Byalistok model the interactions between these two sources of knowledge.

5 Lexical Knowledge: a Common Interest

In addition to the main streams of research presented herein, following Saussure's definition of the syntagmatic and paradigmatic axes [Saussure 49], some linguists have elaborated on this and tried to come up with a formalization of lexical relations [Firth 57], [Chomsky 65], [Halliday 66], [Mel'čuk 81].

In parallel, following Hornby's [Hornby 42] long-standing interest in the study and classification of cooccurrence knowledge ($COOK$), a lot of work has been done in this direction in lexicography and language learning [Cowie 78, 81], [Mackin 78], [Benson 85].

Our aim in this final chapter is to show that although $COOK$ has been investigated from different viewpoints and for different purposes, these investigations can be seen as a common effort to formalize and account for it in dictionaries and linguistic descriptions, and could be successfully used in a computer program.

5.1 A Linguistic Account of Lexical Relations

5.1.1 Early Works

The study of lexis has occupied an important place in linguistics as a field of theoretical investigations since Saussure. Firth in [Firth 57] was the first to recognize that the study of lexis was essential to a full-fledged linguistic model, and that the observation of collocations reveals a lot on both syntactic and semantic levels. Chomsky in [Chomsky 65] proposed a further refinement of the notion of lexical relations. He points out that the reasons why two words co-occur in the same text are not always relevant to a general linguistic description of the considered language. For instance, in the phrase: *to commit suicide*, the cooccurrence of *commit* and *suicide* is clearly of a linguistic type, since *commit* behaves as a syntactic device operating on *suicide*. In contrast, in the phrase: *hold a book*, *hold* could as well be replaced by other related words such as *carry, bring* ..., and the resulting phrase will also be of related meaning. Chomsky termed the two types of cooccurrences respectively, *close constructions* and *loose associations*. Halliday in [Halliday 66] went further, and proposed considering the interactions of lexemes through syntagmatic lexical relations as an independent linguistic level: *lexis*.

For example consider the following two example sentences:

1) "*The ambassador of Freedonia delivered a strong protest concerning the violation of his country's sovereignty.*"

2) "*The ambassador of Freedonia* gave a high *protest concerning the violation of his country's sovereignty.*"

In the first sentence, the fact that *deliver, strong* and *protest* cooccur is not only due to the fact that they have compatible semantic features and that " ... *delivered a strong protest* ..." is grammatical. This is exemplified by the fact that, in the second sentence, "... *gave a high protest* ..." is awkward though grammatical. The difference in well-formedness of these sentences is dependent on the lexical level. This particular type of well-formedness has also been termed "lexicalness" by Halliday in [Halliday 66]. *Deliver, protest* and *strong* are bound by lexical affinities. They are close constructions since, replacing *Ambassador* by *King* or *Emperor* would only change the meaning or the truth value of the sentence, whereas, replacing *strong* by *high* or *big* would change its lexicalness. Such affinities describe lexical cooccurrence knowledge. They embody the knowledge necessary for the proper usage of words and stand for the extent of which an item is specified by its collocational environment.

Mel'čuk in [Mel'čuk 81] contributed to the formalization of close constructions in specifying the different lexical relations. He integrated lexis as part of his complete linguistic model: the Meaning-Text Theory.

5.1.2 The Word According To Mel'čuk

The work of Mel'čuk and his colleagues is geared towards a mathematical formalization of language competence. Natural language in this work is viewed as a logical device mapping all possible texts to all possible meanings. This account of language competence is restricted in several ways. First, there is no claim of cognitive modeling; that is, although the task is to account for human performance, the way it is formalized is not representative of psychological activities. Situational and referential aspects of language are also voluntarily left out. The basic goal of the theory elaborated by Mel'čuk and his colleagues at Moscow and Montreal is to establish a bidirectional mapping between any given meaning and all the texts that express it.

The theory is thus named Meaning-Text Theory (*MTT*). *MTT* draws on other linguistic theories such as stratificational grammars [Lamb 64]. The mapping between meaning and texts is done in seven levels/stages. Going from one level to its immediate successor/predecessor is precisely described by a set of passage rules. At each stage of the mapping, the lexicon is a central source of knowledge, and the formalization of *COOK* is done through the notion of lexical function (*LF*).

5.1.3 *LF*, *COOK* and the Lexicon

In *MTT*, the notion of *LF* is crucial for the description of both paradigmatic and syntagmatic affinities between lexemes. A lexical function (*LF*), is a mapping between words (or phrases) and sets of words (or phrases). Each *LF* stands for an abstracted lexical relations into which words can enter. In other words:

Let $\mathcal{W}$ be the universe of words, and $\mathcal{L}$ the set of *LF*s:

$$\forall (w_1, w_2) \in \mathcal{W},\ \forall \lambda \in \mathcal{L}, \lambda : \mathcal{W} \rightarrow \wp(\mathcal{W})$$
$$\forall x^i \in \lambda(w_i)\ where\ i\ \in \{1, 2\},$$
$$x^1\ and\ w_1\ enter\ into\ the\ same\ relation\ as\ x^2\ and\ w_2$$

Around sixty language-independent primitive *LF*s have been identified. They can be composed and result in other (less primitive) *LF*s. Three examples derived from [Mel'čuk 73] are:

Magn: Associates nouns to adjectives (or adverbs). The role of Magn(< *noun* >), is to emphasize, magnify or stress the meaning of < *noun* > when used in combination with it.

Common examples are:
Magn*(escape) = [narrow, ...]*
Magn*(argument) = [strong, vigorous, courageous, ...]*
Magn*(tea) = [strong, ...]*
Magn*(car) = [powerful, ...]*
Magn*(corpus) = [large, ...]*
Magn*(sound) = [loud, ...]*

Oper: Associates nouns to verbs. Any element of Oper($< noun >$), is a semantically empty verb that takes $< noun >$ as its direct object. The verb is here (and in the next *LF* also) a syntactic device, operating on the $< noun >$. Common examples are:
Oper*(attention) = [pay, ...]*
Oper*(lecture) = [give, ...]*
Oper*(alarm) = [set, ..]*
Oper*(subpoena) = [serve, issue, ...]*

Labor: Associates nouns to verbs. Labor is very similar to Oper, however, the $< noun >$ and its Oper value enter into a different kind of syntactic relation than the $< noun >$ and its Labor value. Here the semantically emptied verb takes $< noun >$ as its indirect object, the indirection being specified in the entry. Common examples are:
Labor*(esteem) = [hold (hold someone into esteem), ...],*
Labor*(consideration) = [take (take something into consideration), ...].*
Labor*(arrest) = [place (place somebody under arrest), ...].*

Lexical functions capture a very important aspect of lexicalness, and numerous examples could be given for the three above *LF*s [Benson 86]. The value of these functions (such as *hold, place,* ...) is strongly language specific and is not predictable in terms of syntax or semantics; however, the functions themselves represent a very important part of language independent knowledge. Abstracting and classifying lexical relations into which words can enter is a very important step forward. The key idea being that, beside the grammatically constrained *closed class words*,[13] a lot of words we utter are mainly used for structural reasons, *i.e.*, their presence is somehow specified/required by their environment. Closed class words are dealt with in the grammar and open class words are dealt with in lexis. The difference is not only a question of quantity but also a question of quality. Grammar deals only with syntactic classes of elements whereas lexis directly deals with the words.

In previous approaches, the lexicon was a remote source of knowledge. It was mainly used to store syntactic (or semantic) features or a-typical morphology. Moreover, its structure was often non-existent or arbitrary. In contrast, *MTT* focuses on the lexicon. In its framework a typical lexical entry consists of several independent zones of linguistic knowledge; one of them, the lexical-cooccurrence zone describes *COOK*. Each lexical entry has a slot for each *LF* in its lexical cooccurrence zone. Ideally, in the process of producing (or understanding) a sentence, these zones are used in order to map a meaning to a sentence bearing it (or conversely).

[13] Closed class words are words belonging to small syntactic categories, such as articles, modal verbs etc. They are opposed to open class words such as nouns, or adjectives that contain numerous words [Huddleston 85].

5.2 Lexicography, Language Learning, and Lexis

There is a huge gap between the formal description of a language and the set of rules[14] that is actually learned by second language learners. This gap cannot be bridged without an immersion in terms of linguistic exposure to $L2$. Given a standard dictionary and a description of the grammar of $L2$, no human (nor a computer) can ever learn $L2$. There are a lot of linguistic phenomena that have not been formalized to the same extent as grammar rules have been. Examples of such phenomena are:

- The "vouvoiement"[15] in French (and its German or Spanish counterpart), is very difficult to master for second language learners and seems to be very hard to explain in a textbook.
- There are 8 genders in Wolof (the main language in Senegal) and the way to use them is very intricate. Their mastery comes with practice. The same thing can be said of genders in some languages such as French or Hebrew that disagree on many words (things). For example, *window* is feminine in French while it is masculine in Hebrew. In Wolof it is even more intricate since genders don't even follow a "universal" designation, such as feminine, masculine or neutral. For example: fruits that grow on trees are grouped into the same gender category, words designating members of the family (such as father, sister, etc.) also belong to one gender.
- The way Israelis give the time in modern Hebrew often violates the rules of standard Hebrew, and yet there is a striking agreement among them on the way to do it, depending on the exact time of day. It has been shown [Shanon 84] that native speakers of modern Hebrew though unaware of the rules specifying it, all agree on when to violate standard Hebrew agreement rules. A linguist willing to formalize this phenomena, would have problems since there is no abstraction or generalization clearly visible.

All those examples fall under the same kind of linguistic phenomena, where people manifest behavior in accordance with the rules of language and are unaware of the rules they apply. Moreover, if formalizable, those rules are often not formalized. They know "how" but not always "what". In general we can say that there is a gap between the set of rules already formalizing the language learned and the set of rules learnable by a second language learner. Part of this gap between the grammar books and the set of rules describing the language can be bridged by $COOK$. Its role in the framework of language learning is highlighted by the behavior of second language learners [Leed 79], an example will illustrate this:

Having a standard dictionary of English and a good grammar book, a second language learner of English whose first language is Hebrew or French will face the following problems while trying to use the words *dream* or *attention*:

[14] By rule, we intend the customary meaning of the word.

[15] The French "vouvoiement" is a politeness device. When addressing persons of a higher social rank, the form "vous" (second person plural) is rather than the familiar "tu" (second person singular).

Instead of saying: *I pay attention to ...*, if the first language is French s/he would say: *I make attention at ...* [*"Je fais attention à..."*]. Similarly, if the first language is Hebrew, instead of *I had a dream ...* s/he would say: *I dreamt a dream ...* [*"h'alamti h'alom ..."*]. Such examples of production by second language learners are numerous and account for a significant part of second language learners errors.

Following the incentive of Hornby ({*e.g* [Hornby 42, Hornby 52, Hornby 65] to compile dictionaries accounting for $COOK$, a lot of work has been done at the intersection of the fields of lexicography and language learning (also called pedagogical lexicography). This work has resulted in a further formalization of $COOK$ and in several dictionaries of English (and in English) for second language learners [Cowie 78], [Hornby 42].

Dictionary entries are not limited any more to the syntactic and semantic definition of the lexeme, but also contain collocational knowledge [Mackin 78]. This knowledge provides the benefit of a huge linguistic expertise that is of great help to the second language learner willing to produce in $L2$. This work has also influenced lexicographers mainly interested in dictionaries for native speakers. Drawing on this work, Benson in [Benson 85] used a subset of the LF described in [Mel'čuk 73] as a framework for the lexicographic description of English. His work can be seen as an effort to remedy the weak structure of $COOK$ in previous lexicographic works.

One of the side-effects of these works has been to extend the work of lexicographers. They now have to study large corpora of English texts of all nature in order to extract and compile $COOK$. This particularly overwhelming task is often carried on through the careful study of large samples of English texts, the study of other dictionaries, the own competence of the lexicographer or the intensive testing of native-speaker's competence [Mackin 78]. In an attempt to relieve lexicographers from the burden of collecting and classifying occurrences, Choueka in [Choueka 83], has proposed algorithms that allow automatic scanning of large textual corpora (millions of words) and to retrieve frequent idiomatic and collocational expressions. Although more interested in the retrieval of commonly used expressions such as: *United Nations, Middle East, Home run, President Reagan, etc.*, his work describes an interesting methodology for handling large corpora and can be considered as a first step toward automated lexicography.

6 Conclusion: Simulation and Isolation

Whether it be for producing or understanding English sentences, processing natural language can be seen as the cooperation of several active components, each one pivoting around and manipulating a particular source of language-related knowledge. Synthesizing the work presented herein, we can say that a complete model of natural language processing must account for the following sources of language-related knowledge[16]:

Abstract Syntactic Knowledge

[16]We only take into consideration here, the processing of single-sentence texts, and neglect the pragmatic aspects of language use.

	FOUL-UP	CHILD	RINA	BERWICK	DEBILI
LANGUAGE RELATED KNOWLEDGE:					
Abstract Syntax	Innate	Innate	Innate Acquired	Innate Acquired	Innate
Intensional Semantics					
Cooccurrence Knowledge					Acquired
Extensional Semantics	Innate Acquired	Innate Acquired	Innate Acquired		
Memory Chunks (SPG)	Innate	Innate	Innate		
World Inferences			Innate		
Psychological Knowledge		Innate			
INPUT DATA	set of sentences	dialogue	dialogue	set of sentences	texts
PERFORMANCE TASK	SAM	itself	itself	itself	disambi-guation
LINGUISTIC MODEL	Schank	Schank	Schank Wilensky	Chomsky	Saussure
COGNITIVE TASK	L1 acquisition	L1 acquisition	L2 acquisition	L1 acquisition	none
PSYCHOLINGUISTIC MODEL	none	none	none	Chomsky	

Figure 16: Evaluation table of the five programs

This knowledge is actually composed of: the specific syntax of the language, and the way to use it for producing or understanding sentences. This knowledge has been compiled and successfully used in computer applications under several different formalisms.

Lexical Knowledge
A Lexicon must be used, giving information on:

- Syntactic properties of the lexeme, such as its syntactic category, government pattern, etc.
- Intensional semantic properties, defined as internal relations between the lexeme and other lexical units. This semantic information is to be distinguished from the encyclopedic information used in the conceptual dependency model.
- Cooccurrence knowledge, the syntagmatic environment of the lexeme is described in term of close constructions.

It should be noted that abstract syntactic and lexical knowledge together constitute what we call linguistic knowledge.

Encyclopedic knowledge. The problems inherent to encyclopedia, have led artificial intelligence researchers to separate it into three different sources:

- Referential word meaning. This is the world counterpart of the intensional meaning; here the semantic information is of denotational type.
- Memory chunks representing specific mundane events. Scripts, plans, goals and lately Memory Organization Packets (*MOPs*) [Schank 82], represent a variety of frozen events stored in memory.
- Inference rules allowing reasoning about the previous memory chunks.

Acquiring a language actually means acquiring the three kinds of knowledge cited above, as well as proficiency in their use. In a computer program such as the ones described in this paper, language-related knowledge (whether learned or innate) is restricted to one or two of the above. For instance, RINA's innate knowledge consists of the knowledge of: abstract syntax, the encyclopedia (in its three forms); and RINA learns referential knowledge of phrases. This information is compiled in figure 16, for the five systems we reviewed in this paper. Each column represents a program and the first seven rows specify the language related knowledge that is learned or innate. The other rows give information on the programs such as: the cognitive processes modeled, the linguistic model used, etc. One of these is the entry for the performance task used for these learning programs. Debili and Granger are the only two having a performance task other than running the program itself on a larger set of sentences; Granger aimed at giving more robustness to SAM, and Debili tackled the problem of disambiguation which is still open. In general we believe that a performance task should be used to evaluate language learning programs, as it is for any other learning program.

For language learning programs, determining the innate and learned knowledge is also a way to decide on the orientation of the program; that is, whether it is a cognitive model, an evidence for a theory, or simply a way to compile linguistic knowledge that is needed for other purposes. Berwick's motivations are to demonstrate the learnability of syntax and thus confirm theoretical assumptions; Zernik, Granger and Selfridge are more interested in the cognitive aspects of the problem; Debili wanted to disambiguate French sentences and was in need of the $COOK$ that he acquired. Aside from disambiguation, $COOK$ has been shown to be of general use in a linguistic model, and its automatic acquisition would relieve lexicographers from a particularly overwhelming task while being of use in general natural language systems as well [Smadja 88].

In conclusion, we can say that among the three kinds of knowledge considered by the programs reviewed in this paper, $COOK$ seems to offer the most advantages. $COOK$ is easily testable in various performance tasks (such as language generation, lexicography, disambiguation, ...) although $COOK$ is not yet represented in machine usable format. Its acquisition is an important task faced by both lexicographers and language learners, and its automatic acquisition and use is now possible thanks to the formalization introduced by linguists. For a computer program, it would constitute a challenging as well as a provide useful results.

7 Acknowledgements

I would like to thank Kathy McKeown for her help during the writing of this paper, she had the courage to go through it lots of times. Many thanks are also due to Gerald Maguire Jr. and Salvatore Stolfo for their insightful comments.

References:

[Adjemian 76] C. Adjemian, *On the Nature of Interlanguage System.* Language Learning 26, pp: 297-320, 1976.

[Allen 80] J.F. Allen & C.R. Perrault, *Analysing Intentions In Utterances.* Artificial Intelligence 15-1, pp: 143-178, January 1980.

[Atkinson 82] M. Atkinson, *Explanations in the Study of Child Language Acquisition.* Cambridge study in Linguistics, Cambridge University Press, 1982.

[Bailey 74] N. Bailey, C. Madden & S.D. Krashen, *Is there a natural sequence in adult second language acquisition?* Language Learning 24-2, pp: 235-243, 1974.

[Baker 82] L.L. Baker & J. A. MacCarthy, *The Logical Problem of Language Acquisition.* MIT Press, Cambridge MA, 1982.

[Barwise 83] J. Barwise & J. Perry, *Bradford Book: Situation And Attitudes.* The MIT Press, Cambridge Ma, 1983.

[Benson 85] M. Benson, *Collocations and Idioms.* In *Dictionaries, Lexicography and Language Learning*, R. Ilson (eds.), Pergamon Institute of English, Oxford, GB, 1985.

[Berwick 85] R.C. Berwick, *The Acquisition of Syntactic Knowledge.* MIT Press, 1985, Cambridge, MA.

[Berwick 86] R. Berwick, in R.S. Michalsky, J.G. Carbonell & T.M Mitchell, *Machine Learning Part Two.* Morgan Kaufmann 1986.

[Berwick 87] R.C. Berwick & S. Pilato, *Learning syntax by automata induction.* Machine learning 2:9-38, 1987. Kluwer Academic Publishers, Boston - Manufactured in the Netherlands.

[Bresnan 82] J. Bresnan & R. Kaplan, *Lexical Functional Grammar.* In J. Bresnan (eds.) , *The Mental Representation of Grammatical Relations.* MIT Press, Cambridge MA, 1982.

[Boyer 85] M. Boyer & G. Lapalme, *generating paraphrases from meaning-text semantic networks.* Computational Intelligence, vol. 1, no. 3-4, 1985.

[Bowerman 78] M. Bowerman, *The Acquisition of word-meanings: an investigation into some current conflicts.* In N. Waterson & C. Snow (eds.), *The development of Communication.* Chichester, 1978.

[Byalistok 78] E. Byalistok, *A theoretical Model of Second Language Learning.* Language Learning 28-1, pp: 69-83, 1978.

[Carey 78] S. Carey, *The Child as a Word Learner.* in M. Halle, J. Bresnan & J. Miller *Linguistic Theory and Psychological Reality.* MIT Press, Cambridge, MA.

[**Carey 83**] S. Carey, *Constraints on the meanings of Natural Kind Terms.* in T.B. Seiler & W. Wannenmacher, *Concept Development and the Development of Word Meaning.* Springer series in Language and Communication 12, Springer-Verlag, New-York Tokyo 1983.

[**Choueka 83**] Y. Choueka, T. Klein and E. Newitz. *Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus.* ALLC journal, vol. 4, 1983. pp. 34-38.

[**Chomsky 57**] N. Chomsky, *Syntactic Structures.* Mouton, The Hague, 1957.

[**Chomsky 65**] N. Chomsky, *Aspects of The Theory of Syntax.* MIT Press, 1965.

[**Chomsky 75**] N. Chomsky, *Reflections on Language.* NY Pantheon Books.

[**Chomsky 81**] N. Chomsky, *Lectures on Government and Binding.* Dordrecht: Foris 1981.

[**Chomsky 86**] N. Chomsky, *Knowledge of Language: Its Nature, Origins and Use.* Praeger, New York NY, 1986.

[**Clark 74**] E. Clark, *Some Aspects of the Conceptual Basis for First Language Acquisition.* Papers and Reports on Child Language Development, Stanford University 7: 23-51.

[**Clark 77**] H. Clark & E. Clark, *Psychology and Language: An Introduction to Psycholinguistics.* NY, Harcourt Brace Jovanovitch, 1971.

[**Comrie 81**] B. Comrie, *Language Universals and Linguistic Typology.* Chicago University Press, Chicago 1981.

[**Cowie 78**] A.P. Cowie, *The place of illustrative material and collocations in the design of a learner dictionary.* In *In Honour of A.S. Hornby.* Oxford University Press, Oxford England, 1978.

[**Cowie 81**] A.P. Cowie, *The Treatment of Collocations and Idioms in Learner's dictionaries.* Applied Linguistics, 2-3 (1981), pp. 223-235.

[**Cruse 86**] D.A. Cruse, *Lexical Semantics.* Cambridge University Press, 1986.

[**Cullingford 78**] R. Cullingford, *Script Application: Computer Understanding of Newspaper Stories.* PhD Thesis, Yale University, Department of Computer Science, TR-116-1978.

[**Debili 82**] F. Debili, *Analyse Syntactico-Semantique Fondée sur une Acquisition Automatique de Relations Lexicales-Sémantiques. Thèse de doctorat d' état*, Paris XI University 1982, Orsay, France.

[**Ducrot 79**] O. Ducrot & T. Todorov, *Encyclopedic Dictionary of the Sciences of Language.* The John Hopkins University Press, 1979.

[**Dyer 83**] M.G. Dyer, *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* MIT Press, Cambridge MA, 1983.

[**Elhadad 87**] M. Elhadad, *Links Between Situation and Language use, Towards a More Coherent Interaction.*Technical Report, Columbia University, New York NY, December 1987.

[Fillmore 78] C.J. Fillmore, *On the Organisation of Semantic Information in the Lexicon.* In Proceedings of the Chicago Linguistic Society, 1978.

[Firth 57] J.R. Firth, *Modes of Meaning. In Papers in Linguistics 1934-1951.* Oxford University Press.

[Gazdar 85] G. Gazdar, E. Klein, G. Pullum & I. Say, *Generalized Phrase Structure Grammar.* Harvard University Press, Cambridge, MA, 1985.

[Goodman 55] N. Goodman, *Fact, Fiction and Forecast.* Bobbs-Merril, Indianapolis, IN, 1955.

[Greenberg 78] J.H. Greenberg, *Universals of Human Language.* Stanford University Press, Stanford, CA, 1978.

[Granger 77] R.H. Granger, *FOUL-UP A program that figures out meanings of words from context.* Proceedings of the 5th International Conference on Artificial Intelligence, 1977. Morgan Kaufmann Publishers, Los Altos, CA.

[Griffiths 76] P.D. Griffiths, *The ontogenic development of lexical reference.* Unpublished doctoral dissertation. University of Edinburgh.

[Grimshaw 82] J. Grimshaw, *Form Function and the LAD.* in [Baker 82]

[Halliday 66] M.A.K. Halliday, *Lexis as a Linguistic Level.* In C.E. Bazell, J.C. Catford, M.A.K Halliday and R.H. Robins (eds.), *In memory of J.R. Firth* London: Longmans Linguistics Library, 1966, pp: 148-162.

[Halliday 75] M.A.K. Halliday, *Learning How To Mean. Explorations in the Development of Language.* Edward Arnold Ltd, London 1975.

[Halliday 76] M.A.K. Halliday, *Lexical Relations.* In *Halliday: system and function in Language* G.R. Kress (eds.). London: Oxford University Press, 1976.

[Heider 71] E.R Heider, *"Focal" color areas and the development of color names.* Developmental Psychology, 4, 447-455, 1971.

[Hornby 42] A.S. Hornby, E.V. Gatenby & H. Wakefield, *Idiomatic and Syntactic English Dictionary.* Kaitakusha, Tokyo, Japan, 1942.

[Huddleston 84] R. Huddleston, *Introduction to the Grammar of English.* Cambridge Textbooks in Linguistics, Cambridge University Press, 1984.

[Hyams 86] N.M. Hyams, *Language Acquisition and the Theory of Parameters.* D. Reidel, Boston MA, 1986.

[Jackendoff 76] R. Jackendoff, *Toward an Explanatory Semantic Representation.* Linguistic Inquiry 7-2, pp: 59-150, 1976.

[Kay 79] M. Kay, *Functional Grammar,* in Proceedings of the 5th Meeting of the Berkeley Linguistic Society, 1979.

[**Katz 72**] J.J. Katz, *Semantic Theory.* New York: Harper Row, 1972.

[**Katz 80**] J.J. Katz, *Chomsky on meaning.* Language, vol. 56-1, March 80.

[**Keenan 71**] E.L. Keenan, *Two Kinds of Presuppositions in Natural Language.* In C. Fillmore, D.T. Laugendoen, Holt, Reinhart & Winston (eds.) , *Studies in Linguistic Semantics.*

[**Keil 81**] F.C. Keil, *Constraints on Knowledge and Cognitive Development.* Psychological Review, 88, 197-227, 1981.

[**Klein 86**] W. Klein, *Second Language Acquisition.* Cambridge University Press, Cambridge 1986.

[**Krashen 82**] S.D. Krashen, *Principles and Practice in Second Language Acquisition.* Oxford Pergamon Press, 1982.

[**Lamb 66**] S.M. Lamb, *Outline of Stratificational Grammar.* Georgetown University Press, Washington, D.C. 1986.

[**Langley 87**] P. Langley, *Machine Learning and Grammar Induction.* Machine learning 2:5-8, 1987. Kluwer Academic Publishers, Boston, MA, 1987.

[**Leed 79**] R.L. Leed & A.D. Nakhimovsky, *Lexical Functions and Language Learning.* Slavic and East European Journal, Vol. 23-1, 1979.

[**Lyons 68**] J. Lyons, *Introduction to Theoretical Linguistics.* Cambridge University Press, London, England, 1968.

[**Magiste 85**] E. Magiste, *Development of Intra- and Interlingual Interference in Bilinguals.* Journal of Psycholinguistic Research 14-2, 1985.

[**Mackin 78**] R. Mackin, *On collocations: words shall be known by the company they keep.* In *In honour of A.S. Hornby.* Oxford University Press, Oxford, England, 1978.

[**Marcus 80**] M. Marcus, *A Theory of Syntactic Recognition for Natural Language.* MIT Press, Cambridge, MA, 1980.

[**Markowitz 86**] J. Markowitz, T. Ahlswede, M. Evens, *Semantically Significant Patterns in Dictionary Definitions.* Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics, ACL 1986.

[**Mel'čuk 73**] I.A. Mel'čuk *Lexical Functions in Lexicographic Description.* Proceedings of the Berkeley Linguistics Society, 1973.

[**Mel'čuk 81**] I.A Mel'čuk, *Meaning-Text Models: a Recent Trend in Soviet Linguistics.* The annual review of anthropology, 1981.

[**Mitchell 82**] T.M. Mitchell, *Generalization as Search.* Artificial Intelligence 18, pp: 203-226, 1982.

[Morris 71] C.W. Morris, *Approaches to Semiotics: Writings in the General Theory of Signs.* Mouton, The Hague 1971.

[Nelson 79] K. Nelson, *Features. Contrasts and the FCH: Some Comments on Barret's Lexical Development Hypothesis.* Journal of Child Language 6:139-46.

[Quine 60] W.V.O Quine, *Word and Object.* MIT press, Cambridge MA.

[Reich 76] P.A. Reich, *The early acquisition of word meaning.* Journal of Child Language. vol. 3, pp: 117-23, 1976.

[Riesbeck 81] C. Riesbeck, *Conceptual Analysis* in R.C. Schank & C. Riesbeck. *Inside Computer Understanding.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1981.

[Rosch 73] E.H Rosch, *On the internal structure of perceptual and semantic categories.* In T. Moore (eds.), *Cognitive Development and the Acquisition of Language.* New York, Wiley, 1973.

[Rosch 75] E.H Rosch, *Cognitive representations of semantic categories.* Journal of Experimental Psychology, general 104, pp: 102-233, 1975.

[Rutherford 84] W. Rutherford, *Language Universals and Second Language Acquisition.* Benjamins, Amsterdam.

[Rutherford 86] W. Rutherford, *Grammatical Theory and L2 Acquisition a Brief Overview.* Second Language Research, vol. 2-1, June 1986.

[Saussure 49] F. De Saussure, *Cours de Linguistique Générale, 4ème edition.* Librairie Payot, Paris, France, 1949.

[Selfridge 81] M. Selfridge, *A Computer Model of Child Language Acquisition.* Proceedings of the Seventh International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers, Los Altos, CA, 1981.

[Schank 72] R.C Schank, *Conceptual Dependency: A theory of natural language understanding.* Cognitive Psychology, Vol. 3-4, 1972.

[Schank 82] R.C Schank, *Dynamic Memory, A Theory of Reminding and Learning in Computer and People.* Cambridge University Press, New York, 1982.

[Searle 69] J.S. Searle, *Speech Acts, An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge England, 1969.

[Shanon 85] B. Shanon, *How to tell the time in Hebrew ?* Journal of Psycholinguistic Research, vol. 14-3, 1985.

[Smadja 88] F.A. Smadja, *Lexical Cooccurrence: The Missing Link in Language Acquisition.* 15th International Conference of the Association for literary and Linguistic Computing (ALLC), Jerusalem, Israel, June 1988.

[**Vanlehn 87**] K. Vanlehn & W. Ball, *A Version Space Approach to Learning Context-Free Grammars.* Machine Learning vol. 2:39-74. Kluwer Academic Publishers, Boston, 1987.

[**Walters 87**] J. Walters & Y. Wolf, *Second Language Comprehension Processes: Theories and Methods.*, 1987.

[**Wanner 82**] E. Wanner & L. Gleitman, *Language Acquisition: The State of The State of the Art.* In E. Wanner & L. Gleitman,*Language Acquisition: The State of the Art.* Cambridge University Press, Cambridge, 1982.

[**White 83**] L. White, *Markedness and Parameter Setting: Some Implications for a Theory of Adult Second Language Acquisition.* Presented at the 12th Annual University of Wisconsin Linguistic Symposium, March 1983.

[**Wilensky 80**] R. Wilensky, Y. Arens, *PHRAN: A Knowledge Based Approach to Natural Language Analysis.* In Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, 1980.

[**Zernik 87**] U. Zernik, *Strategies in Language Acquisition: Learning Phrases from Examples in Context.* (UCLA-AI-87-1). PhD dissertation, computer science department, University of California, Los Angeles, CA, 1987.