Machine Learning, 16, 87–120 (1994)

Simulating the Child's Acquisition of the Lexicon and Syntax—Experiences with *Babel*

RICK KAZMAN rnkazman@watcgl.uwaterloo.ca
Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Editor: Michael Pazzani

Abstract. This paper describes the theory and implementation of *Babel*, a system which explores the hypothesis that much of the differences in the world's languages may be characterized by the inventory and properties of the lexical items and functional categories of those languages. The structure of *Babel* assumes that functional categories are originally lacking in a child's syntax, and are acquired through a statistical induction process of lexical acquisition. *Babel* then uses information induced from the structure of the lexicon to create a model of syntax via a deductive, rule-based process. This model makes a number of predictions about the time course of language acquisition. These predictions are tested by running *Babel* as a simulation of child language acquisition, using large samples of adult speech to children as input. The simulation results are shown to highly correlate to longitudinal studies of child language acquisition in English and Polish. Finally, the approach to handling noisy data with *Babel* is detailed.

Keywords: statistical induction, government-binding theory

1. Introduction

A large number of machine learning systems have been developed over the past two decades with the goal of automatically acquiring the syntax, lexicon, morphology and semantics of natural language.[1] Some of these systems have had an additional goal: that the acquisition of language should model human language acquisition—learn constructs in the same order and make the same mistakes as a human, utilizing cognitively plausible data structures and algorithms ((Anderson, 1975), (Selfridge, 1980), (Berwick, 1985), and (Rumelhart, 1986) are some of the better known attempts). However, few of these systems have made an explicit goal of demonstrating a statistical correspondence between their results and empirical observations of children acquiring language.

Most acquisition systems have attempted to model the semantic or syntactic stages of acquisition, where a stage is a set of logically related pieces of semantic and syntactic knowledge. However to speak of distinct stages of acquisition is to use an artificially constructed convenience which implies an all-or-nothing state of knowledge on the part of the child. These stages conveniently characterize a child's state of knowledge at a gross level, but do little to account for the detailed time course of acquisition—the mistakes which children make, the regressions, the inconsistent use of a construct or grammatical item.

Because of this characterization of acquisition as a set of clearly defined stages, each of which has some triggering conditions, many of these systems when exposed to a carefully constructed sequence of a few dozen English sentences could learn the majority of English syntax ((Selfridge, 1980) and (Berwick, 1985), for example). These are clearly cognitively implausible models. Furthermore, most language learning systems have used only English as their target language, but any system which purports to model human language acquisition must not be language specific. While it has not been shown that these systems *could not* learn languages other than English, it hasn't been part of the common research agenda to demonstrate their facility with other languages.

(Pinker, 1979), discussing the features that one should demand of a model of human language acquisition, outlined six conditions as criteria by which a formal model of language acquisition might be judged. The model of acquisition described here is intended to meet all of these criteria with respect to a crucially important stage in language acquisition—roughly paralleling an 18 to 30 month old child. This year-long stage is probably the most fecund stage in all of language acquisition, when the bulk of common lexical items and the basic grammatical rules of the language are acquired. The acquisition criteria are as follows:[2]

1. *Coverage*. The model must be able to acquire some sizable fragment of human spoken language.

2. *Input plausibility*. The acquisition model can only make use of information to which the child can reasonably be assumed to have access. So, for instance, the model cannot rely on negative evidence, since there is strong evidence suggesting that children do not make use of it.[3]

3. *Empirical fidelity*. The acquisition model should acquire the phonological, morphological and syntactic structures of language in the same order as children, making the same mistakes as they do.

4. *Time limitations*. The model should not take an excessively long time to acquire the basics of language. In practice, this means that the model must not require, for success, more utterances than an average child would hear in about one year.

5. *Non-specificity*. The model should not be biased towards a single language, but should work equally well with any language, since children appear to learn all spoken languages with equal ease.

6. *Cognitive constraints*. The model should only make use of cognitive abilities that the child can be reasonably assumed to possess. For instance, a model that remembers every sentence ever heard would not be cognitively plausible.

The model of lexical and syntactic language acquisition presented here was designed to provide a formal, computational link between the acquisition of the lexicon (and its inflectional morphology) and syntax (and its agreement phenomena). Informally, *agreement* refers to the morphological changes which can potentially occur when two syntactic structures are related in a sentence, and the phonological changes which audibly signal this relationship.

Recent research in syntax (Abney, 1987), (Fukui, 1987) (Chomsky, 1989) indicates that inflectional agreement morphology and syntactic agreement structures are linked—they are different manifestations of the same types of linguistic information.[4] In empirical studies with *Babel*, a system designed to implement this linkage in a computational model of language acquisition, it turns out that many well-known phenomena of language acquisition—the acquisition of functional categories, agreement, and case marking, the overregularization of affixes, the stages of categorial knowledge—can be accounted for by a unified model of lexical and syntactic acquisition (Kazman, 1991b), (Kazman, 1991c). Furthermore, *Babel*'s results do not depend on language particular assumptions. As will be shown, *Babel*'s mechanisms work across languages.

The link between lexical morphology and syntactic agreement is mediated by the functional categories (FCs) of a language. In fact, much of the syntax of the world's languages may be characterized by the inventory and properties of the lexical items and FCs of those languages. Briefly, FCs are made up of the "little words" of a language: determiners, auxiliaries, complementizers, prepositions and inflectional affixes. They are distinguished, as a class, from the thematic categories (TCs). TCs are made up of the open-class words of a language: verbs, nouns, adjectives and adverbs.[5] I will be investigating the proposal that syntax is acquired by the child as a progression from an invariant base (a core grammar which is common to all languages, and which innately includes TCs) to a more articulated view of language which includes FCs. Given that the inventory of FCs varies widely from language to language, it is necessary for any account of child language acquisition to explain how children come to learn the inventory and properties of a language's FCs.

2. Related work

In order to properly understand the significance and success of the methods presented here, it is illustrative to look at previous attempts at creating computational models of child language acquisition. This section will examine a semantics-based system (Anderson, 1975), a conceptual-dependency based system (Selfridge, 1980), a system based upon Government-Binding (GB) theory (Berwick, 1985) and a connectionist model (Rumelhart, 1986).

2.1. LAS

Anderson's Language Acquisition System (LAS) (Anderson, 1975), is a semantics-based system. It learns languages by attempting to fit semantic representations of input sentences (which it represents in the form of trees) to structural representations of the sentence. This model presupposes that the child understands all of the content words (thematic words) in the input sentence. Given an input of a sentence, accompanied by a semantic tree representation, the system attempts to fit the tree to the sentence, transforming the tree if necessary.

When a part of the input sentence can not fit into the semantic representation, the grammar is modified to account for it. For instance, LAS can rearrange the order of nodes in the semantic tree, so long as the dominance relations between nodes are maintained, and so long as the resulting tree contains no crossed branches. This restriction against crossed nodes is one of LAS's biggest shortcomings. Since LAS has no notion of differing levels of representation for a sentence (such as D- and S-structure in GB theory), it must interpret all sentences relative to the surface order of words. As such, it cannot handle so-called free word order languages such as Walpiri or Polish, since these languages frequently contain discontinuous constituents. In these languages LAS would be forced to postulate crossed branches to join up the various parts of a phrase.

Furthermore, LAS cannot handle constructions with crossed branches even in a relatively fixed word order language such as English. For instance, in the sentence *In the upcoming biathalon, I am going to swim and ride across that lake and over that mountain.*, the prepositional phrases *across that lake* and *over that mountain* can only be attached to their respective verbs *swim* and *ride* by crossing branches. Once again, this relatively unproblematic sentence would be unlearnable by LAS.

LAS made an important contribution to the field at a time when models of language acquisition were weak, underspecified or empirically inadequate. Aside from the restriction of no crossed branches in the input, LAS uses, for the most part, psychologically plausible mechanisms, and makes reasonable assumptions about the form of the input. LAS doesn't remember every sentence presented to it, but rather works on the basis of its current state of knowledge, and refines this in the presence of non-conforming data.

However, LAS is a tremendously powerful system—it can learn large subsets of its target languages after only a dozen or so input sentences—and it is this power which makes it an unrealistic model of human language acquisition. In addition, LAS has a great deal of difficulty learning the function words of a language—determiners, prepositions, case markers, etc. I will argue that the analysis of these elements is not only crucial to the success of a language acquisition model, but, more to the point, the acquisition of these facts determines the very structure of the language to be acquired.

2.2. CHILD

(Selfridge, 1980) has also attempted to implement a semantics-based learner called CHILD, using Conceptual Dependency (CD) graphs as the underlying representation of sentences. CDs are representations of the semantics of a sentence, with the syntactic knowledge represented by a "sequential structure" which "specifies the location of the fillers for slots in CD concepts in CHILD's short-term memory." The process of language acquisition, in the CHILD system is one of learning the appropriate syntactic positions of words which can fill the slots in a CD structure.

Although this model manages to explain more facts of language acquisition in a reasonably natural way, it must eventually be rejected as an accurate view of human language acquisition for several reasons. Like Anderson's LAS, CHILD can learn grammatical facts from a single exposure to the data—something which children show no evidence of doing. Also, CHILD has no mechanism for incorporating closed-class words into its comprehension mechanism, since these play little or no part in the meaning of a sentence. Finally, Selfridge's system has no way of making syntactic generalizations, as children and adults do.

Similarly, CHILD's model of lexical acquisition (Selfridge, 1981) can account for some of the observable facts about the development of a child's lexicon, e.g. the observation that children go through a stage in which they say *goed* rather than *went*. However CHILD accounts for these observations in a relatively *ad hoc* manner: the lexicon is simply ordered as a list, with recently heard entries moved to the front of the list. When the child attempts to use a word, he searches from the front of the list. This representation of the lexicon bears no relationship to current psychological models of the mental lexicon (Emmorey, 1987).

CHILD's architecture is able to mimic many observations of child speech, for example that children typically say *went* initially, then overgeneralize to *goed* and finally learn that *went* is, indeed, the proper form. Selfridge does this by assuming that, initially, *went* precedes *go* in the lexicon, and so is accessed first by the child. At some point, the past tense affix is learned (although Selfridge never clearly states how lexical acquisition proceeds) and placed in the lexicon ahead of *went* and *go*. Next the child learns that *go* + *-d* indicates the past tense of *go* (again, Selfridge is unclear on how this happens) and so produces *goed*, presumably because both *go* and *-d* are still ahead of *went* in the child's lexicon.

Selfridge never explains why the child should eventually fix on *went*, rather than *goed* as the proper form, except to say that "the child will continue to say 'goed' at least until he has heard 'went' again" (Selfridge, 1981). This suggests that the child only needs to hear one example of *went* in order to correct his overgeneralization. This is clearly not the case. The period of overgeneralization in child speech lasts well into the child's 5th year (Marcus, 1990). However, overgeneralization is never common in a child's speech, accounting for only about 2.5% of the child's utterances. Selfridge's model would predict that overgeneralization is an all-or-nothing occurrence.

2.3. Berwick

(Berwick, 1985) has endeavored to create a realistic computational model of language acquisition which is able to acquire a significant subset of English. He does so following the principles of a syntactic theory, and under strict learnability assumptions.

Berwick's model receives input sentences as segmented strings annotated with information such as thematic roles and lexical features. If, at any point, the parse fails with the current grammar, the acquisition phase is entered. The acquisition phase is given, as its state, the instantaneous description (ID) of the parser at the point of failure: the contents of the input buffer, the type of node currently active (Specifier, Head or Complement), the names of previously executed grammar rules, and so on. In response, the acquisition system attempts to modify its base of rules in order to accommodate the new piece of evidence. For instance, the acquisition system tries to attach the input word to the active X' constituent. If this fails, the system attempts to compose a new transformational grammar rule to account for the input. Finally, if any new rule has been created by the acquisition process, a generalization procedure attempts to combine this with some existing rule.

The proposed system achieves a significant amount of success as a psychologically real model of language acquisition. It closely follows the principles outline by Pinker in (Pinker, 1979). It suggests only reasonable assumptions about the cognitive nature of the learner, makes psychologically plausible assumptions, learns the target language (English in this case) quite readily and appears to be reasonably general.

Berwick's system, however, does not actually constitute a realistic model of child language acquisition, because it does not meet the empirical fidelity and time limitation constraints. The system can, for instance, learn rules based upon a single piece of evidence—something which no study of child language has ever indicated a child can do. Furthermore, the system can learn rules in orders which would not be evidenced in actual acquisition studies. These complaints notwithstanding, Berwick's work stands as an important and influential metric against which other proposals for models of language acquisition should be judged.

2.4. *Connectionist models*

More recently, connectionist models have begun to be investigated (Rumelhart, 1986). Given that the Rumelhart and McClelland (henceforth R & M) model has been the most widely discussed and cited connectionist model of language learning in recent years, it is important to examine the claims made both for and against it. R & M claim that a connectionist model of language may be able to do away with the notion of *rule* altogether, and that all of a human's knowledge of language may be represented in a connectionist network. To demonstrate this claim, they provide a model of the acquisition of English past tense facts.

Furthermore, the R & M model, once trained, is able to account for a surprisingly large number of facts of English language acquisition—facts on which the model was never explicitly trained. For instance, the model is able to learn English regular and irregular past tense rules. It initially learns irregular and regular verbs correctly, then overregularizes (e.g. *getted, throwed*), and finally achieves mastery of both regular and irregular verbs, thus exhibiting the so-called "U-shaped" learning curve which has been claimed for lexical acquisition in children. R & M's model can achieve this with no explicit representation for a word, a morpheme or a rule.

While the goals and achievements of this model are substantial, R & M's model has been shown to have serious difficulties, both methodological and practical (among its critics are Pinker & Prince (Pinker, 1988) and Lachter & Bever (Lachter, 1988)). Its critics find that the connectionist model does no better, and sometimes worse, than rule-based models of language acquisition in terms of its account of the facts of language acquisition. Furthermore, Rumelhart and McClelland's system does not solve the language induction problem, contrary to their claim, since it both overgeneralizes and undergeneralizes inappropriately.

Pinker and Prince delineate twelve major problems with R & M's model, which will only be mentioned here briefly (for the details, the reader is encouraged to examine (Pinker, 1988)). These problems are: the ways in which R & M's model accounts for the regularization of the past tense morpheme is incorrect; the model's successes have nothing to do with its Connectionist architecture (and hence can be easily duplicated in symbolic models, such as *Babel*); it makes many false predictions about morphology and phonology and can represent rules not found in any human language, and yet is not powerful enough to represent all words in English; it makes many errors; and it cannot account for psychological similarity and distinctions among words.

In addition to these criticisms, the success of the empirical results of R & M's model appears to crucially depend upon a special order of data presentation. One feature of their model is that it mimics the "U-shaped" learning curve which children are supposed to exhibit in the acquisition of past tense endings. However, in order to achieve this result, Rumelhart and McClelland had to resort to a special order of data presentation. This presentation of data renders the results achieved by their model highly suspect.

In the first stage, the model received 10 iterations of the same 10 verbs—2 regular and 8 irregular. In the next sample, it received 190 iterations of 410 different verbs, approximately 80% of which were regular—exactly reversing the proportions of regular and irregular verbs in the input, as compared with stage 1. At this point, the model begins to overregularize. Given this enormous change in the size and composition of the input in the second stage, the model reacts by overextending the use of the regular form, because the regular form is dominant in the input. Thus, the U-shaped curve in the R & M model is merely an artifact of the statistical patterns in the data which is fed to it. Furthermore, this pattern of data presentation is not substantiated by studies of the input to children (Pinker, 1988). Finally, even the premise that children's acquisition of the past tense follows a U-shaped curve has been strongly challenged recently (Marcus, 1990), thus calling into question the validity of a computational model which mimics this phenomenon.

3. Outline of *Babel*'s acquisition model

Babel's approach to modelling child language acquisition combines some of the features of the approaches outlined in section 2: it uses an explicit theory of grammar (GB theory) as a basis for explaining what a child knows about syntax. *Babel* attempts to motivate changes in the child's state of syntactic knowledge (represented as the rules of GB theory) by connecting that state of knowledge to generalizations made in the lexicon.

The analysis of the lexicon and the development of syntactic structures appear to be significantly different, although not unrelated, classes of problems. Accordingly, different learning strategies are applied to them: *Babel* employs an inductive approach to the acquisition of the lexicon and a deductive approach to the acquisition of syntax. Thus, the knowledge of the lexicon is accrued through a process which shares much with the connectionist models. The lexicon is constructed inductively, driven by regularities in the input data. This is accomplished through a statistical ranking procedure which postulates hypotheses for words—essentially candidate lexical entries—based upon the frequency, phonological salience, and semantic and syntactic features of input words. The main purpose of the lexical acquisition procedure is to discover the root word forms and the productive affixes of a language.

Once the inductive procedure has discovered the form and meaning of the productive affixes of a language, the structure of the lexicon can be said to be acquired. Certainly, there is a great deal more to lexical acquisition than this. For example, this proposal is only concerned with inflectional morphology. Derivational morphology is a complex system which the child must also learn, possibly independently. However, once the *structure* of the lexicon is learned, much of lexical acquisition is reduced to a relatively rote process of adding new entries which conform to previously learned paradigms.

That is, once the acquisition procedure has learned the verbs *walks, throws, wants, gives*, etc., learning *brings* or *takes* involves only the application of a rule that the child has already learned, i.e. add the appropriately phonologically conditioned form of the *-s* affix to the verb root to produce the 3rd person singular present tense form.[6]

The candidate lexical entries which the lexical acquisition procedure proposes for word roots and affixes are ranked, and their ranks are continually updated based upon how well they account for input. In this way, productive roots and affixes will be reinforced by the input, whereas less productive forms will either be pruned or simply listed in the lexicon as irregulars. Productive inflectional affixes encode agreement information on their host category—for instance, the verbal *-ed* affix in English encodes past tense. This piece of information about *-ed* cannot be innate however. It represents an inductive generalization on the child's part, based upon exposure to a large number of regular English verbs.

The deductive, syntax-learning portion of the model is logically dependent upon the results produced by the inductive, morphology-learning portion. The knowledge base for making deductions about the syntactic structure of the target language is the lexicon. The lexicon, when properly structured, contains detailed information about the agreement properties of the categories of a language—agreement is manifested in the affixes which attach to lexical roots. This dependence upon the inductive lexical acquisition process can be seen as a way of buffering a deductive model of syntax from the enormous amount of variation and inconsistency in language. Once the inductive analysis has provided a clear model of the lexicon, the deductive portion of the model may safely make generalizations based upon this information.

The deductive portion of the model is structured as a production system of if-then rules associated with a database of facts, where any rule may be activated when its preconditions are met in the database. To give a concrete example, we might postulate the following rules as part of the deductive model of syntax:

IF a syntactic category X exhibits regular agreement properties
THEN hypothesize an Agr(eement) node dominating X″

IF a syntactic category X has an Agr node and a function word exists in the lexicon which subcategorizes for X
THEN analyze this function word in X″s Agr node

Rules such as these provide a means to gradually refine the target grammar based upon an analysis of grammatical relations revealed in the lexicon. In this paper, we only examine categorial projection—that is, how a child comes to understand the representation and properties of a category in his language—and so these two rules are the

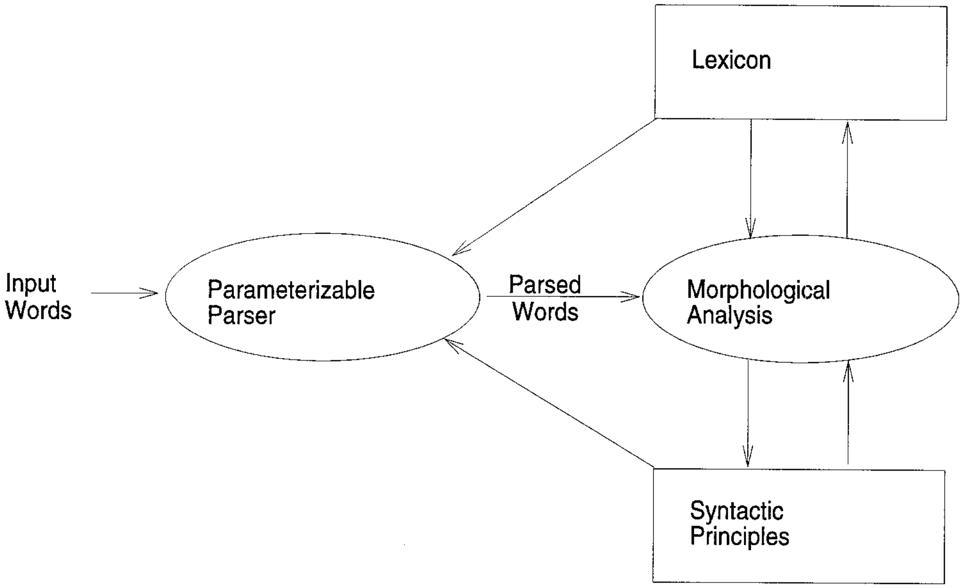

Figure 1. The Architecture of *Babel*

only ones which are used. These particular rules determine if and when two categories should be constrained to agree in the syntax. Such rules are necessary because children begin their linguistic careers with no knowledge of agreement and gradually develop this knowledge over time. Different languages exhibit different agreement facts, and so agreement information must be deduced by the child as part of the acquisition process. Part of *Babel*'s contribution, then, is to propose a way in which this knowledge might be acquired by the child. *Babel*'s success should then be judged on how closely the acquisition of agreement parallels a child's acquisition of agreement, and whether *Babel* utilizes psychologically plausible mechanisms to achieve its ends.

The architecture of *Babel* is shown in Figure 1 (in this diagram, ovals are processes, rectangles are repositories and arrows represent flow of information). The input to the parser is assumed to be phonetic forms of words, along with a semantic representation (which indicates the meaning of each word, along with any syntactic attributes that it possesses, e.g. number, tense, mood, definiteness). The assumption of paired phonological and semantic representations as input to the acquisition mechanism is common among language acquisition researchers (Anderson, 1974), (Wexler, 1980), (Grimshaw, 1981), (Pinker, 1984), (Berwick, 1985). It has considerable empirical support as well. For example, it has been shown (Slobin, 1975) that children are exposed to very few sentences which they cannot decipher through contextual clues.

The morphological analysis mechanism logically stands at the center of the entire model of acquisition, accepting input from a parser (in the form of the words of fully parsed sentences, annotated with information gleaned from the parsing process) and updating the lexicon and the knowledge base of syntactic principles. The parser is assumed to be a parameterizable parser, which encodes basic notions such as X′ theory, government, case assignment, the θ-criterion, and so on. Examples of a parser of this sort may be found in (Fong, 1989) and (Gibson, 1991). These two processes—parser and morphological analysis—and two repositories—lexicon and syntactic principles—have been found adequate to model a wide range of acquisition processes.

4. *Babel*'s algorithm

A pseudo-code version of *Babel*'s language acquisition algorithm is presented in Figure 2. The input to tests of *Babel* is the spontaneous speech of adults talking to children, as recorded in the CHILDES database (MacWhinney, 1985). The sample which we will present in detail is taken from English language transcripts. Although 10,165 different types were found in the data sample, the first 1,000 of these types (those with a frequency of at least 50 uses in the entire sample) accounted for almost 91% of the words spoken. These 1,000 types were the raw material from which the large data set was prepared.

From these 1,000 types, representing over 665,000 tokens, the following types of words were removed: all pronouns, exclamations ("hey", "okay", "yeah", "ta"), all forms of the copula, contractions, negators ("don't", "no", "not", "won't") verb-particle idioms ("used to", "supposed to") and proper names. Some of these, such as proper names, were removed because some were irrelevant to the lexical acquisition process being studied. Other tokens, like contractions, were removed because they involved acquisition processes which were not being modelled, such as the acquisition of phonological clitics. Still other tokens, such as copulas were removed because their acquisition involves complex syntactic interactions which are not being modelled currently.

The resulting data set of 850 words represented over 442,000 tokens. 10% of this data set statistically balanced according to the frequencies of input tokens in the full data set, was used as input to *Babel* producing the results given in Section 7.

A sample of the input data is given in figure 3. Each input word is given in phonetic form (enclosed in square brackets), using an ascii-based phonetic alphabet. The notation used for the phonetic form was derived from a dictionary used by the Carnegie Mellon University Speech Group, and all input words were translated using that dictionary. This is a pure phonetic form, giving only phoneme representations and their stress level (for English, only two stress levels were used, high stress, indicated by a 1 following a phoneme, or low stress, the default). No information about morpheme boundaries is included in the input; these boundaries must be inferred by the acquisition mechanism.

In addition to the phonetic form, a number of other features distinguish each input word: POS is the part of speech, SEM is a unique semantic label (distinguishing the word from all other unrelated words, including synonyms), ARGS are the arguments of the word, IFEATS are the internal features and EFEATS are the external features. A word's arguments are the syntactic structures which it requires in order to create a well-formed syntactic structure. For example, we can say *The dog devoured the meat*, but not **The dog devoured.* This is because the verb *devour* is transitive—it contains

```
for each [word $W_j$]
    if [$W_j$ is not in the lexicon]
    then
        assign an initial structure to $W_j$
        add $W_j$ to the lexicon
    else
        update $W_j$'s lexical entry
        for each [root $R_{s_i}$ and affix $A_{t_j}$]
            increase the rank for any root $R_{s_i}$ which produces the
                proper surface form $W_j$
            increase the rank for any affix $A_{t_j}$ which produces the
                proper surface form $W_j$
            increase the rank of the link between $R_{s_i}$ and $A_{t_j}$[0.2cm]
        if [rank($W_j$) $>$ $\tau$]
        then
            if [$W_j$ is semantically related to some other
                lexical entries $W_{k_1} \ldots W_{k_m}$]
            then
                create a root $R_j$ containing the common phonetic and
                    semantic content of $W_j$, $W_{k_1} \ldots W_{k_m}$
                create an affix $A_j$ which expresses the phonetic difference
                    between $W_j$ and $R_j$
                add $A_j$ to the set of affixes (if it doesn't exist)
                generalize $R_j$ with other affixes which encode the
                    same phonetic change
                attach all deductively acceptable affixes to $R_j$
                initialize affix ranks based on their input frequencies
        if [$W_j$ has an affix $A_i$ and rank($A_i$) $>$ $\rho$]
            project a representation for $W_j$ which includes agreement
```

Figure 2. A Pseudo-code Representation of *Babel*'s Algorithm

two argument positions, a subject (in this case *The dog*) and a direct object (in this case *the meat*) which must obligatorily be filled for the sentence to be grammatical.

Internal features are features which indicate a word's meaning in context. For example, in the sentence *I can touch the forks*, the word *forks* has a set of internal features which

```
[WORD:[aa1 iy] POS:noun SEM:i
        IFEATS:[TH=AGT PER=1 NUM=SG DEF=Y]
        EFEATS:[ROLE=SUB] ]
[WORD:[k ae1 n] POS:aux SEM:can
        ARGS:[[SUB=VERB]]
        IFEATS:[TNS=PRS MOOD=SUBJ] ]
[WORD:[t ah1 ch] POS:verb SEM:touch
        ARGS:[[TH=AGT PER=1 NUM=SG DEF=Y ROLE=SUB]
              [TH=PAT PER=3 NUM=PL DEF=Y ROLE=OBJ]]
        IFEATS:[TNS=PRS MOOD=SUBJ] ]
[WORD:[dh ah] POS:det SEM:the
        IFEATS:[TH=PAT PER=3 NUM=PL DEF=Y]
        EFEATS:[ROLE=OBJ]
        ARGS:[[SUB=NOUN]] ]
[WORD:[f ao1 r k s] POS:noun SEM:fork
        IFEATS:[TH=PAT PER=3 NUM=PL DEF=Y]
        EFEATS:[ROLE=OBJ] ]
```

Figure 3. Sample Input to *Babel's* Lexical Acquisition Component: *I can touch the forks*

are inherent in its meaning (such as plural number, NUM=PL), and internal features which adhere to the word by virtue of its context (such as being definite DEF=Y).

External features, on the other hand, are an expression of the particular relationships which this word undergoes in a given context. The word *forks*, in figure 3 has a single attribute in its set of external features—ROLE=OBJ—indicating that *forks* is acting as an object in this sentence. However the attribute ROLE=OBJ has nothing to do with the inherent meaning of the word *forks*, it is simply a piece of contextual information. Each of ARGS, IFEATS and EFEATS are optional in any given word.

For a complete explanation of the contents and meanings of a *Babel* lexical entry, see appendix A2.

Although we have made the assumption of paired phonological and semantic representations, it is clearly implausible to assume that the child innately and perfectly understands each feature of a sentence, and the word to which that feature belongs. Note that in figure 3 features are shared among phrases, i.e. the input is not a 1:1 mapping between words and features. For example, the words in the noun phrase *the forks* share the same set of internal features (TH=PAT PER=3 NUM=PL DEF=Y) and external features (ROLE=OBJ). It is left to the lexical acquisition process to determine which of these features are inherent in the meanings of the individual words, and which are superfluous. In section 9, a technique for dealing with the more realistic assumption of noisy data is discussed—data where appropriate features may be missing and spurious features may be present on any given input word.

5. Functional categories

The speech of children in the early stages of language acquisition is characterized by a lack of tense, agreement, articles, proper negation and case marking—all of which are FCs. This is exemplified by (1), a sample of data from a child, Nina, at age 2;1 (Suppes, 1973). Surprisingly, in a short period of time, all of these phenomena begin to be productive in the child's spontaneous speech, as shown in (2), a sample of data from Nina at age 2;3—just two months later.

(1a) Three ball.
(1b) Fall down. Falling down.
(1c) You jamas. You jamas.
(1d) My have more.

(2a) I will give him a book.
(2b) I don't have valentine.
(2c) Nina's Daddy.
(2d) It was too wet in your playroom.

At this point, one might ask: why should the acquisition of agreement and independent function words have anything to do with each other? Language researchers have posited a number of compelling arguments in favor of treating function words and inflectional affixes as a single class (Abney, 1987), (Fukui, 1987).

The unified treatment of FCs is justified on five grounds: 1) they have similar syntactic characteristics—for instance, FCs form a distinct class with respect to restrictions on movement (Baker, 1990); 2) FCs have similar morphological characteristics—they are generally unstressed, dependent morphemes (typically clitics or affixes) (Abney, 1987); 3) they have similar semantic characteristics—they have no independent reference, but rather modify their hosts (which are TCs) meaning; 4) in Broca's aphasics, FCs are lost as a group (Caplan, 1988); and 5) FCs are acquired as a group by children (Kazman, 1990).

By explicitly linking the acquisition of agreement and function words in a model of language acquisition, not only are we able to make significant and subtle generalizations about the target language, but we are able to closely model the time course of language acquisition cross-linguistically. This model makes a number of predictions about the time course of acquisition which have been shown to hold for English, Polish, Dutch, French and Hebrew (Kazman, 1991a), (Kazman, 1991d). Only the results for English and Polish will be discussed in this paper.

In each of these languages, it has been shown that the rate with which children learn inflectional affixes is directly proportional to the frequency and phonological salience of those affixes in the input, and that the inflectional affixes for a category are always acquired by the child just before he acquires the independent function words of that category.

6. Lexical acquisition

The morphological analysis component of *Babel* was designed to provide a principled account of certain observations noted by language acquisition researchers:

- Children do not initially morphologically analyze words (Brown, 1973), (DeVilliers, 1985).
- Children overgeneralize inflectional morphology (Ervin, 1964), (Clark, 1985), (Berman, 1985). They create over-regularized forms like *taked*, and occasionally create doubly marked forms, like *ated* and *feets*. Although children typically overgeneralize in only about 2.5% of their spontaneous word productions (Marcus, 1990), these overgeneralizations must still be accounted for; they are a key piece of evidence indicating the child's state of linguistic knowledge.
- The rate of acquisition of lexical items is directly proportional to the input frequency and phonological salience of these items (Scarborough, 1977), (Berman, 1985), (Clahsen, 1988).
- Children waffle in their acquisition of morphology (as well as syntax). For instance, a child will often produce correct and incorrect plural forms of the same word in succession (Marcus, 1990).
- The lexicon is organized around (possibly abstract) root forms (Emmorey, 1987).

Babel's lexical acquisition procedure discovers the roots and affixes of a language. It does this by comparing semantically related words[7] and creating affixes to describe the phonological and morpho-syntactic attribute changes between related input forms. In the earliest stages of lexical acquisition, the lexicon does not contain enough information to be able to make productive generalizations. The lexicon is simply a set of unrelated words, as depicted in Figure 4.

This representation of the lexicon corresponds to a period when the child shows no knowledge of inflection, and uses words as unanalyzed wholes. When the child has acquired enough samples of the language to be able to make generalizations, then semantically related forms of words are compared and roots and affixes are created (Ervin, 1964).

For example, this procedure would compare *throw* and *throws* and create an affix consisting of the phonological difference, *-s* paired with a set of attribute changes: third person, singular, present tense. The set of features of the affix is determined by taking the set difference of the features of the inflected form of the word and its root. The root is determined by taking the intersection of the phonological and morpho-syntactic features of all related forms of a word (this process is described in detail in (Kazman,

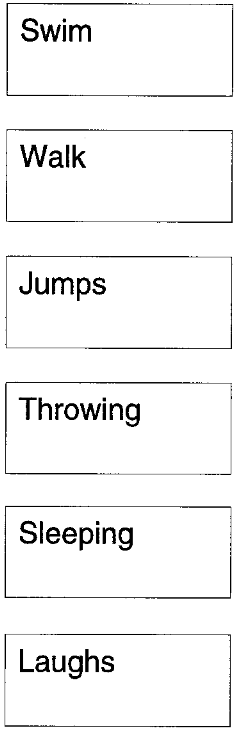

Figure 4. An Unanalyzed Lexicon

1991d)). This procedure of taking the intersection of the phonological and morpho-syntactic features of input words and derived affixes is done continually by *Babel* in the process of learning a language's lexicon. In this way, a lexical entry for a root or affix is refined to contain only those features which are central to the word or affix's representation.

Once an affix such as *-s* has been created, it is free to combine with any word of the appropriate category, if that word provides the correct phonological environment. For example, the *-s* and *-d* affixes must agree in voicing with the phonological material to which they immediately attach. The phonological environment is determined through a process of taking the intersection of the distinctive features (Chomsky, 1968) of the phonological forms of the words to which a particular affix applies. Once affixes have been created, the lexicon has the structure exemplified in Figure 5. In this structure, affixes are included as distinct but bound lexical elements. Surface forms of words are realized by linking a root with an affix.

Affixes, roots and the links between them are all ranked in this model. Different ranking strategies have been employed. For clean input data, the rank of a root, affix or link can be easily computed with a sigmoid function based upon the input frequency of

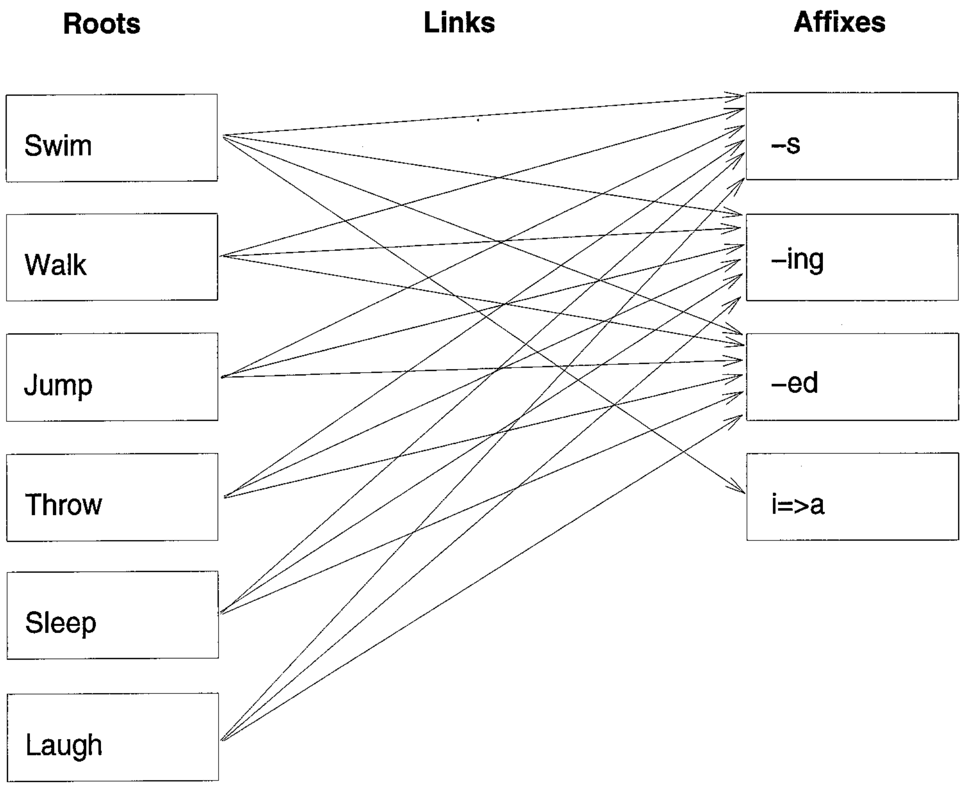

Figure 5. A Partially Analyzed Lexicon

the word: $rank = 2/(1 + (1/e^{usage}) - 1$. For the more realistic assumption of noisy data, a relative-frequency function proves to be more robust, as will be shown in Section 9.

As distinct lexical elements, affixes can compete for use among the words of the language. *Babel*'s lexical acquisition procedure reinforces—increases the rank—of any affix which is utilized in an input word. However, the lexicon is not simply a naive statistical book-keeper, containing all and only those combinations seen in the input. Affixes, once created, are free to associate with other roots in the lexicon. Roots are linked with all deductively acceptable affixes to create the surface forms of words.[8] Thus, in Figure 5 the i→a affix can only attach to the verbal root *swim*, because *swim* contains the appropriate phonological environment—an internal *i*, whereas the *-ed* affix can attach to any verb (but no nouns or adjectives).

The rank of a surface form of a word can be defined as follows: the rank of $word_{ij}$ is defined to be the product of the ranks of $root_i$, the k affixes $(affix_1 \ldots affix_k)$ and the k links $(link_{ij1} \ldots link_{ijk})$ which join the affixes to their root:

$$\text{rank}(\text{word}_{ij}) = \text{rank}(\text{root}_i) * \prod_{h=1}^{k} (\text{rank}(\text{link}_{ijh}) * \text{rank}(\text{affix}_{jh}))$$

For English, k = 1, since roots in English take only a single inflectional affix. This is not the case cross-linguistically however. Note also that this formulation explicitly allows the possibility of having many competing lexical entries for the same word. There may be *i* different roots, each of which have *k* links to *j* affixes.

The exhaustive attachment of appropriate roots and affixes also creates words which are not valid in the adult language, such as *throwed, swimmed* and *sleeped*, but which have been widely attested in the speech of children in the 18-30 month age range. With exposure to input, however, overgeneralizations are corrected. This is because words like *swimmed* and *throwed* are seldom or never heard in the input, and thus the ranks of the links from *swim* and *throw* to *-ed* are not likely to be reinforced. The ability to create forms such as *swimmed* and *throwed*, is retained in the lexical entry forever, but at an extremely low probability. This representation of the lexicon corresponds with empirical evidence—the probability of overgeneralization approaches 0 in adults but never reaches 0 (recent estimates (Marcus, 1990) put the rate in adults at roughly .00004).

With sufficient exposure to input, the productive affixes of the language will eventually be identified, because they will be more quickly reinforced than unproductive affixes. For example, the past-tense *-ed* affix attaches to a great number of verbs in English. However, the internal vowel change past-tense affix $i \rightarrow$a is only found in a few verb pairs such as *swim/swam, ring/rang, sink/sank*, and does not apply in general to other wise appropriate phonological environments (compare *link/linked, skim/skimmed, sting/stung*, and so would only apply to a small subset of lexicon. For this reason, the i→a affix would be reinforced by the input more slowly than the *-ed* affix.

One way of viewing the lexical acquisition process just described is that it is a way of distilling information contained in the lexicon into meaningful classes: the roots and productive affixes of each of the lexical categories. This has two important consequences: 1) the affixes on a TC signal that category's syntactic agreement relations; 2) if we adopt the additional assumption that a category's agreement information has an independent instantiation in syntax, as has been argued for theoretic and cross-linguistic reasons in (Abney, 1987), and for developmental reasons in (Kazman, 1990) then this *agreement* position provides exactly the environment necessary to analyze and incorporate the function words of a language. In this way, the idiosyncratic syntactic structure of a language, as expressed by the language's inventory of functional categories, may be determined through an examination of the properties of the lexicon.

The lexical acquisition procedure of ranking roots, affixes and their links is independently necessary in order to account for the productivity of affixes, for morphological doublets (e.g. *sublet/subletted*) and for psycholinguistic effects, such as memory priming (Emmorey, 1987). In addition, this acquisition procedure carries with it a number of logical consequences which can be empirically verified:

1. That children will overgeneralize regular affixes, but seldom irregulars (Ervin, 1964);
2. That the frequency of overgeneralizations of any given word, and of the lexicon in general, will asymptotically decrease over time (as has been demonstrated in (Marcus, 1990));
3. That noisy data will retard, but not disrupt the process of acquisition;
4. That children will waffle in their use of words, i.e., use a correct and an incorrect form of the same word in alternation.

To explain point 1: some affixes, like *-en*, which turns the singular *ox* into the plural *oxen* are almost never heard in the input, and so would have extremely low ranks (the rank of *-en* would only be updated on the rare occasions when *oxen* is heard in the input). The cognitive architecture manifested in *Babel* would predict that affixes such as *-en* are virtually never overgeneralized. This is exactly what is attested in acquisition studies (Ervin, 1964). Point 2 has been demonstrated by Marcus *et al* (Marcus, 1990), as has been discussed, and point 3—the robustness of *Babel* in the face of noisy data—will be discussed in Section 9. Finally, the waffling in word production—point 4—can be understood if we interpret the ranks of words as their relative probabilities of use in language production.

7. Syntactic acquisition

Now that the lexical acquisition model has been described, the assumptions and workings of the syntactic acquisition model must be made explicit. The parser which the model utilizes shares all lexical features among the words within a phrase (this is the notion of percolation of features, found in (Emonds, 1985), (Davis, 1991) and elsewhere). Furthermore, phrases may attach to other phrases as either arguments or adjuncts.

Parses of input are made according to the child's current model of grammar (syntactic principles) and his current lexicon. These syntactic principles and lexicon are both created by and used by *Babel* as shown in Figure 1.

In addition to these assumptions, *Babel*'s assumptions about the child's initial state of knowledge must be made explicit. These are as follows:

- The child understands, and can utilize the mechanisms of *core grammar*, as described in appendix A1, including c-command, government, X′ theory, the θ-criterion and agreement. The child's model of his language may not, however, have enough detail to actually make use of these mechanisms. For example, while the child innately has the capacity to understand agreement, he does not innately possess any information about what agrees with what in a given language (which is why children say things like *three cookie*). Agreement particulars must be acquired through the inductive lexical acquisition process described in the previous section.
- The child's initial parsing model utilizes a minimal X′ representations, with thematic words projecting to simple bare phrases, and no representation at all for function

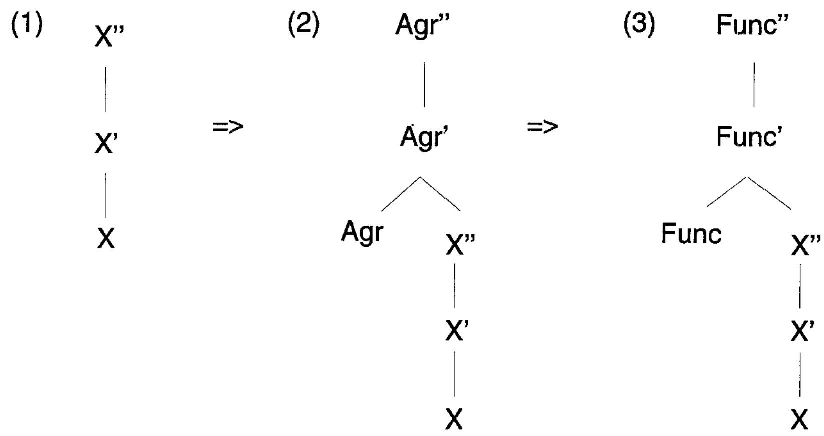

Figure 6. Three stages of categorial representation

words or agreement. This corresponds to the child's earliest attempts at speech production where function words, if they are used at all, are used indiscriminately, e.g.

Nina: On the my dress.
I want see it.
Nina reading.
Make a crying the baby.

- It is assumed that the child's earliest parses are representations of simple predication relationships. Words which are not subjects or predicates (typically function words and adjuncts) will be analyzed as affixes or will be syntactically adjoined to their appropriate heads. This is why children say *want ball* and not *I want the ball.*

With these assumptions, the child can parse virtually any input he receives, although the resulting parses will not be assigned the same structure that an adult would create. This is as it should be. The child's grammar, as evidenced by sentences (1a-d) in section 5, indicates that the child is using a vastly different grammar from the adult, lacking agreement and function words. These differences can be accounted for with the assumption that the child initially projects all TCs identically according to X′ theory and has no representation for FCs, as in projection 1 of Figure 6 (Kazman, 1990).

By analyzing the agreement properties of each lexical category, as evidenced by their affixes, the child will gradually learn which categories exhibit regular agreement processes: predictable meaning changes paired with changes in the phonetic form of a category. These are precisely the categories which have productive affixes. For these categories, the child will posit an agreement node (Agr), dominating the lexical category, as in projection 2 of Figure 6. As this stage, the child possesses the mechanisms to say *John sleeps now*, rather than *John sleep now*.

Finally, if the child hears function words which are manifestations of a thematic category's agreement features, then these words will be identified with that category's Agr position, as indicated in projection 3 of Figure 6, by the re-labelling of the Agr position as Func. At this stage the child is able to parse sentences such as: *John can sleep now*. The use of projections 2-3 are exemplified by Nina's speech in samples (2a-d). This stage of grammar more closely resembles the adult grammar. Nina uses agreement, tense and independent function words, although her speech is not perfectly adult-like.

Although these constructions are proposed as models of the child's development, they are not *ad hoc*. Each of the representations 1-3 of Figure 6 is a valid X′ representation for adult language. This can be stated confidently because examples of each kind of construction exist in the languages of the world: there are lexical categories which make use of no agreement or function words (Adjectives in English, or any category in Chinese or Japanese), which would be represented as 1. There are categories which exhibit agreement but contain no function words (nouns in Polish or Russian, verbs in Quechua), which would be represented as 2. Finally, there are categories which both exhibit agreement and contain function words (like nouns and verbs in English or any Romance language), which would be represented as 3 (Abney, 1987). Furthermore, these stages of categorial representation mirror empirical studies of language acquisition, as shown in the studies cited herein and in (Kazman, 1991d).

By allowing the input to dictate which thematic categories will be simple projections of the head (as in 1), which ones will exhibit syntactic agreement but no function words (as in 2), and which ones will contain function words (as in 3), a model of the language can slowly be built by the acquisition process. This method provides a way to tailor a maximally general initial grammar so that it can adequately represent the idiosyncrasies of a particular target language. Furthermore, as the child builds more complex, more accurate representations for the categories of his language, he can begin to correctly parse more of his language, allowing still more complex representations to be entertained. In this way, the lexical acquisition procedure provides the foundation for the long process of syntactic bootstrapping.

Babel is able to produce accurate predictions of child language acquisition given no prior information about a language: it is able to learn both classes of words and affixes, and syntactic constructions in the same order that children do. *Babel* learns the lexicon of a language by employing the following procedure: *Babel*'s parser is fed adult speech to children as its input, and outputs annotated X′ structures as parses. The parsed words become the input to the lexical acquisition procedure which determines the roots and affixes of the language. The hypothesized affixes are exhaustively attached to all appropriate root forms in parallel, and the roots, affixes and links between the roots and the affixes are ranked.

When *Babel*'s lexical acquisition mechanism receives a new piece of input, it does two things: 1) it exhaustively updates the ranks of all roots, affixes and links which can correctly account for the input form, and 2) it attempts to further add to or refine its knowledge of the roots and affixes in its lexicon by comparing this input word to previously learned related words. When a root form is attached to an affix which is productive (above an arbitrary threshold τ), *Babel* hypothesizes a complex syntactic projection for this word which includes Agr(eement). This productivity threshold τ was set to 0.9 (out of 1.0) in tests of *Babel*. This value was chosen because it parallels the notion of acquisition which Brown (Brown, 1973) and the de Villiers (DeVilliers, 1985) used in their empirical studies of child speech: 90% correct usage in obligatory contexts. Finally, if *Babel* finds function words which subcategorize for a particular category, and that category has an Agr projection, a Func node will be hypothesized in place of the Agr node. Both agreement affixes and independent function words may then reside in Func.

8. Simulation results

The results of a *Babel* simulation (using the 44,000 word input sample of speech of care-givers talking to children), are shown in table 1, which compares the order of acquisition of 8 different English grammatical morphemes predicted by this model, with the order in which children have been shown to acquire these morphemes in two seminal empirical studies: Brown's (Brown, 1973), and the de Villiers' (DeVilliers, 1985). The morphemes tested in these studies include: affixes such as regular past tense (*-ed*), progressive aspect (*-ing*) and the plural (*-s*); independent function words such as articles (*a/the*) and prepositions (*in/on*); and clitics such as the possessive (*'s*). In order to determine the time of acquisition transcripts of children's spontaneous speech were examined in (Brown, 1973) and (DeVilliers, 1985). Once the morpheme in question was being used by the children correctly 90% of the time in obligatory context, the morpheme was said to be acquired, thus producing the orderings shown in table 1.

The relative order of acquisition for *Babel* was determined by a similar method. The ranks of the roots and affixes in the lexicon produced by *Babel*, along with the syntactic structures which had been hypothesized, were inspected at regular intervals (every 1,000 input words) in the processing of the 44,000 word input sample. When a morpheme had acquired a rank of 0.9, it was determined to have been "acquired" by *Babel*. For example, *Babel* would, after some exposure to English input data, discover the plural morpheme *-s* in all three of its phonological variants (found in, for example, *books*, *oranges* and *gloves*), along with the correct attributes which characterize this affix, namely NUM=PL (plural number). Since this affix is highly productive in English, its rank is increased relatively quickly.

In table 1, the plural morpheme was assigned a rank of 4 by *Babel*, 4 by Brown, and 2 by the de Villiers.[9] This means that, for *Babel*, it was the fourth morpheme to achieve a rank of 0.9 in the 44,000 word test run, whereas for Brown and the de Villiers, it was the 4th and 2nd morpheme, respectively, which achieved a level of 90% correct usage in obligatory contexts among the children they studied.

The Pearson product moment coefficient of correlation was calculated for the three sets of results. The order of acquisition of the 8 grammatical morphemes produced in this test of *Babel* is correlated at a level of 0.99 with Brown's results, and 0.96 with the de Villiers' results.

Table 1.
Experimental Results: Brown vs. *Babel* vs. de Villiers

Morpheme	**Order of Acquisition**		
	Brown	*Babel*	**de Villiers**
Present Progressive	1	1	2
on	2	2	2
in	2	3	4
Plural	4	4	2
Possessive	5	6	6
Articles	6	5	5
Past Regular	7	7	7
3rd Person Regular	8	8	7

Babel has also been tested (Kazman, 1991d) with a relatively small set (1,000 words) of Polish data. Although the data set was nowhere near large enough to establish a statistically significant experiment (and similarly, highly detailed studies of the order of morphemic acquisition such as the Brown and de Villiers studies do not exist for Polish) the results do show that in a highly inflected language such as Polish the rates of acquisition of inflectional paradigms were much higher than languages like English or French, which have little inflectional morphology.

Since the evidence for inflection on nouns, verbs and adjectives in Polish was so abundant in the input and so phonologically salient (affixes in Polish are always syllabic), the ranks for affixes in the test of *Babel* grew far more quickly than affixes in English. As a result, this model would predict an unusually fast acquisition of affixes in Polish, as compared with English, and a correspondingly rapid hypothesis of complex categories for nouns, adjectives and verbs in Polish. This is exactly the pattern which is found in Polish language acquisition studies (Weist, 1984), (Smoczynska, 1985): Polish children, after initially using a fixed word order and uninflected word forms, quickly move to marking gender, tense and aspect by about age 2;0, and simultaneously move to freer word order. Furthermore, by this age they make relatively few errors in their use of inflectional affixes. In English, children of age 2;0 frequently make inflection errors. This pattern is exactly what would be predicted by a model such as the current one, which links the acquisition of inflection and functional categories to the frequency, regularity and salience of inflection in the input.

The test of Polish served two other functions:

1. It demonstrated that the acquisition mechanisms in *Babel* are not language specific. There is no special processing used to analyze English which could not, in principle, be used to learn Tamil, Spanish or Walpiri. The only language particular code in the system is the module which translates words from their phonetic representations into the internal distinctive feature representation. Once the input has been translated into its internal representation, all languages are treated identically.

2. It showed that *Babel* could create root forms of words which were total abstractions. Unlike English, a bare root form (say, of a verb) is never expressed in Polish. So, although a child in English might hear *play* in its root form (e.g., *Marsha and Maggie*

play), a child in Polish will only hear verbs in their inflected forms—the root form of a word in Polish is an abstraction. However, Polish language acquisition researchers (Smoczynska, 1985) claim that Polish children do know the abstract root forms of words, on the basis of children's spontaneous novel forms.

9. Noisy data

Virtually all of the implementation issues discussed thus far have been under the assumption of an idealized input stream. Although this assumption provides an excellent background for examining the theoretical consequences of a model of acquisition, it is clearly untenable for a model which claims to be psychologically valid. The input which a child receives is noisy. The types of noise which the language learner might encounter are things like: dialectal variation, contact with other children (whose utterances do not necessarily conform to the adult model), contact with non-native speakers of the language, as well as imperfect hearing or comprehension of an utterance or its context. In each of these cases, the input which the acquisition device receives will be incorrect in some manner.

Examples of input cited heretofore have included all and only those attributes which are relevant to a particular lexical item and the phrase which immediately contains it. This assumes perfect understanding of an utterance directed toward the child. Such an assumption is clearly untenable in practice. Thus, even though it has been assumed that children get, as input, paired semantic and phonological representations of the input, the child may not have detailed understanding of the correspondences between the overall semantics of an utterance and the semantics of any given word in the utterance. This section, then, is devoted to exploring a technique for filtering the "noisy" input which a child will receive in the real world, in order to arrive at a representation which more closely reflects his true knowledge of language. That such a process exists is undeniable—the knowledge of language possessed by a population of speakers of the same dialect is quite homogenous. However, the form of this filtering process is a matter for investigation.

The ways in which an incorrect piece of input data can be represented in this model are as follows: the input item could have extra, spurious attributes, it could be missing crucial attributes, or it could have incorrect attributes which are substituted for the true attributes of the word. Furthermore, the input item's phonetic representation could be rendered incorrectly.

As an example of an additional attribute, the child might hear the word *jump*, and believe that it encodes third person singular agreement. This might occur through a simple misunderstanding on the child's part, or as a result of hearing a dialectal variation of English where the third person singular form does not differ from the first or second person, i.e., *He jump over the box*. As a consequence, in his input, the child would represent the word jump as follows:

[WORD:[jh ah1 m p] POS:verb SEM:jump
    ARGS:[[TH=AGT **PER=3** NUM=SG DEF=Y ROLE=SUB]
        [TH=GOAL ROLE=OBJ]]

IFEATS:[TNS=PRS MOOD=IND]]

Similarly, the child might overlook an attribute of a word. If he hears an elliptic sentence such as *That's doggy's*, he might understand the utterance to mean something like *That's a doggy* (this would be a particularly easy mistake if the child had no representation for function words such as the determiner *a*, and the possessive *'s*). In this case, the child would misunderstand the fact that this utterance intended to demonstrate something about ownership. As such, the child's representation of the lexical item *doggy's* would be missing the possessive attribute (POSS=Y):

[WORD:[d ao1 g iy z] POS:noun SEM:dog
IFEATS:[PER=3 NUM=SG DEF=Y DIM=Y]
EFEATS:[ROLE=OBJ]]

Finally, the child might substitute one attribute for another—a correct attribute for an incorrect one. Consider, for instance, the case where the child does not realize that a certain word pertained to the past tense, and nothing in the context of a conversation indicates such to him. For example, if someone said to a child: *The Phoenicians always travelled in large boats*. Unless the child knew who and what the Phoenicians were—i.e., a semitic tribe which ruled large parts of the Mediterranean 2 to 2 1/2 millennia ago—he would have no *contextual* clues as to the tense of this sentence. As such, he might think that the Phoenicians are people like the Browns next door, and that *travelled* refers to the present tense (cf. *The Browns always travel in a large van*). The child's representation for such a version of the word *travelled* as input would then be as follows:

[WORD:[t r ae1 v ah l d] POS:verb SEM:travel
ARGS:[[TH=AGT PER=3 NUM=PL DEF=Y ROLE=SUB]]
IFEATS:[**TNS=PRS** MOOD=IND]]

Instances of these sorts of misunderstandings can be easily found in child language data (MacWhinney, 1985). In each of the above cases, the proper syntactic attributes are not being reinforced by the acquisition procedure. A simple approach to dealing with such types of noise suggests itself from the overall philosophy of this implementation: the syntactic attributes of a lexical item may be individually ranked, just as the roots and affixes which contain these attributes are ranked. These attributes then compete with each other, to represent the true attributes of the input word. The ranks of particular attributes on a given word are then increased when *Babel* detects the use of these attributes in input examples of this word. Rather than updating ranks through a sigmoid function, as was suggested in Section 6, the rank of an attribute F on a given word W is simply the percentage of input tokens of W in which the attribute F appears.

A percentage-based function is employed for ranking attributes because the sigmoid function is an asymptotically increasing function. Such a behavior is untenable as a characterization of attribute ranks in a model with noisy input because if enough examples of an "incorrect" attribute are heard by the child, the rank of this attribute will eventually exceed ρ (the rank level used in other parts of the model to indicate adult-like comprehension by the child). What is really required of the rank of an attribute is not a function of the raw frequency of that attribute, but a function of its *relative* frequency. Thus the percentage of times a given attribute is used, as compared with its alternatives, is its rank. For example, the attributes NUM=PL and NUM=SG are alternatives of the number attribute in English.

By ranking attributes in this way the input is, as in the rest of the model, dictating the nature of the linguistic generalizations made. The inductive lexical acquisition procedure exists solely to note generalizations in the input. A process such as this will be able to determine, for example, that phrases which include the definite article *the* have something to do with definiteness (as represented by the attribute DEF=Y). Since noisy attributes are exceptional—not part of the core meaning of a word—they will eventually be excluded from the word's lexical entry by any process which notes attribute generalizations.

Given this framework for handling noise one problem poses itself. There are many attributes that get attached to a word as part of the process of parsing which are not part of the core meaning of that word. For instance, a noun would often be marked with a definiteness attribute (which it gets through percolation from its determiner), but definiteness is not an intrinsic attribute of most nouns. As a consequence, a noun like *giraffe* will have spurious attributes such as its definiteness or person dictated by its environment within a sentence and these attributes will be regularly reinforced, according to the input, but these spurious attributes will not be part of the core meaning of *giraffe*. The resulting noun would look as follows (where the relative rank of an attribute is indicated, in parentheses, after the attribute):

sem=giraffe rank=0.500520 # args=0 pos=Noun projection=Noun″
phon=jh ah r ae f #
ifeats=[TH=AGT (0.545) PER=3 (0.545) **NUM=SG (0.955)**
DEF=Y (0.545) POSS=N (0.818) TH=PAT (0.364)
PER=1 (0.273) DEF=N (0.455) TH=THM (0.091)
PER=2 (0.182) POSS=Y (0.182) NUM=PL (0.045)]
efeats=[ROLE=SUB (0.364) ROLE=OBJ (0.364) ROLE=2OBJ (0.273)]

This lexical item was achieved by exposing *Babel* to 80 tokens for *giraffe*, 5% of which had noise introduced to them. Each of the noisy lexical items contained exactly one incorrect attribute. Note that the attribute NUM=SG, highlighted in this representation—the only attribute which is truly intrinsic to the noun *giraffe*—has the highest rank of all of the word's attributes. This is expected, given the architecture of the model and the assumption that the child understands most of what he hears. This assumption is shared by virtually all language acquisition researchers and modellers (Anderson, 1975), (Wexler, 1980), (Berwick, 1985), (Pinker, 1988). Without it, it is difficult to provide a foundation for any acquisition of language at all. Note that the incorrect attribute NUM=PL is also included in the representation for *giraffe*, although at a very low rank. This is a reflection of noise in the child's input. Furthermore, various spurious attributes—POSS=N, DEF=Y—are included in this lexical entry with relatively high ranks, even though they are not inherent in the meaning of *giraffe*.

Clearly, some means of distinguishing the core attributes of a word is needed. Given that lexical entries in this acquisition model contain ranked attributes, what metrics can be used to indicate which are the true attributes of a word and which attributes are artifacts of noise or simply the context in which a word is often found? One possibility which has been implemented in *Babel* is to define a level σ, and declare that all attributes which are within σ of the most highly ranked attribute are considered to be the true attributes of the root or affix.

It was empirically determined that a value of 0.1 for σ allowed *Babel* to operate robustly with noisy input—that is, it could learn the true attributes of words even with an input stream containing levels of noise up to about 22%. This is thus a prediction which *Babel* makes: raising σ to higher levels resulted in higher levels of allowable noise, but permitted incorrect attributes to be included in the lexical entry as valid; lower levels of σ did not permit *Babel* to operate under levels of noise as high as 22%. The correct value of noise which children can withstand and still learn their target language is an open empirical question, but it is likely to be much less than 22%. It seems highly unlikely that roughly one in four input tokens to the child will be flawed in some way, particularly given that adults often intentionally simplify their speech to children.

Now that a method for discovering the highly ranked attributes of a lexical entry has been described, there still remains the question of what to do with attributes of intermediate rank. For the majority of attributes which have ranks less than σ below the top-ranked attribute (as is the case for definiteness, person and other attributes in the example of *giraffe*) there are two logical possibilities: these related attributes might be incidental to the meaning of the word, and may all be safely pruned, or these attributes may be indicative of uses of this word in different dialects. Consider, for example, what would happen if the child was in close contact with another child who regularly used the possessive pronoun *my* as a nominative pronoun, as Nina, the child in the Suppes study (Suppes, 1973), does:

Nina: My see it.
My need her.
Let my do it.
My moving the legs.

If a child was constantly exposed to language such as this (a sibling, for example), the child might split the lexical entry for *my* into two parts—one entry which indicated that it was a possessive pronoun, and one entry which indicated that, under certain circumstances (i.e., when talking to Nina), *my* could be a nominative pronoun as well. By storing all of the attributes of a word and ranking them, this process is greatly facilitated. If all but the most highly ranked attributes were discarded, however, this information would be lost and the lexical acquisition procedure would have no way of representing information about specialized dialects—information which is part of everyone's knowledge of language.

On one level, this strategy for dealing with noisy data can simply be seen as a modification of the attribute intersection technique described in section 4. Using that technique we could prune conflicting attributes with impunity since, given the assumption of clean data, conflicting attributes could not be part of the word's intrinsic meaning. The technique being presented in this section achieves the same end by pruning attributes which are more than σ below the most highly ranked attribute—an indication that these low ranked attributes do not show up in the majority of input samples of this word, and so must not be essential to the true characterization of the word.

However, because of the different potential sources of errors it is not possible to know how to set a universally applicable σ for the sorts of variation introduced by factors such as language contact. An appropriate level for σ can only be determined through empirical study. At the present time, no studies exist which systematically characterize the types and frequencies of errors to which children are exposed, and describe the nature of the learning process in the face of large numbers of errors. What is needed is a set of studies which compare groups of children learning language in different environments in—homogeneous populations and in heterogeneous populations.

These complications notwithstanding, the notion of ranked attributes presented here still has intrinsic value: it provides a robust mechanism to implement the screening of noise. Once there are empirical studies which can precisely characterize noise in the child's input the results of these studies may be reflected in the implementation of *Babel*. Furthermore, ranking attributes is a way of capturing the notion of the child's uncertainty with respect to word meaning. Since the child is not likely to make a quantum leap from ignorance to understanding, the notion of ranked attributes can simply and elegantly model the child's growing knowledge of the true characteristics of a lexical item.

10. Conclusions

A proper view of the structure of the lexicon, and the application of different learning strategies—inductive learning for the lexicon and deductive learning for syntax—provides the foundation for the model presented here. These techniques are not, however, simply being proposed for methodological reasons; they are independently justified by longitudinal studies of acquisition and psycholinguistic studies on the structure of the mental lexicon. The lexical acquisition procedure not only learns the productive affixes of the target languages, but does so in a way which is consistent with the acquisition of the lexicon by children: words are first learned unanalyzed, then some words are overgeneralized and gradually, over time, an adult-like knowledge of the lexicon is learned. Words may be misapplied, by misunderstanding an attribute, but such misunderstandings will eventually be retracted.

The analysis of the lexicon provides a means by which the child can analyze function words in his grammar: when the child sees that a category utilizes a set of productive affixes, he instantiates an agreement node for these affixes in the syntax. This corresponds to observations made by language acquisition researchers cross-linguistically: that function words become productive at the same time as, or immediately before children begin to master inflection. This characterization of lexical and syntactic acquisition allows *Babel* to accurately mirror the acquisition of syntax and morphology in children.

Finally, the approach to handling noise in *Babel* is a straightforward extrapolation of the techniques used for lexical acquisition: the syntactic attributes of a lexical item are individually ranked, just as the roots and affixes which contain these attributes are ranked. In this way *Babel* can produce correct results (as compared with behavior of children, determined through longitudinal studies) in the face of noise levels up to 22%. By demonstrating that *Babel*'s results are invariant in the presence of high levels of noise, the validity of *Babel* as a representation of a realistic model of acquisition is strengthened. *Babel* can work not only in the "clean-room" environment of the researcher's lab, but in the "messy-room" environment where we all learn language.

Acknowledgements

I would like to thank Ted Gibson, Michael Pazzani, and two anonymous *Machine Learning* reviewers for their insightful comments. They have improved this paper greatly.

Appendix A1

Government-binding theory basics

Research into Government-Binding theory, sometimes called "Generative grammar", "Transformational grammar", or the "Principles and Parameters" approach to language is a large and complex body of knowledge. In an appendix such as this, the major concepts can be enumerated, but there is no way of properly justifying these concepts, or showing their detailed interactions. For this sort of background, the reader is encouraged to consult with an introductory text, such as (Radford, 1988), and the references cited therein.

One of the objectives of modern theories of grammar is to define a *core grammar*—a set of linguistic principles which characterize the full range of universal linguistic phenomena found in natural languages. The part of a language which is described by these core, or *unmarked*, rules and structures is differentiated from the part of a language which is described by the periphery, or *marked* rules and structures. All languages combine both aspects, but it is the inventory of peripheral rules and structures which causes diversity among languages. This paper explores the hypothesis (pursued by many researchers in recent work on linguistic theory) that the periphery is completely describable as differences in the lexicons of different languages.

In this appendix the following concepts, which are crucial to an understanding of core grammar as it is currently formulated, will be defined: X′ theory, specifiers, complements, adjuncts, C-command, government, the θ-criterion, D-structure, S-structure, Move-α, case marking and agreement. These concepts do not comprise a complete inventory of the principles of core grammar, but illustrate enough of core grammar to support the arguments being put forth in this paper.

X′ theory was motivated by the observation that a language's lexical categories formed phrases in identical ways, and could be described by the general structure in figure A1.1 (where *X* stands for any lexical category and items enclosed in parentheses are optional):

For example, compare:

- the noun *student* with the noun phrase *a student of history*;
- the adjective *good* with the adjectival phrase *very good at chess*;

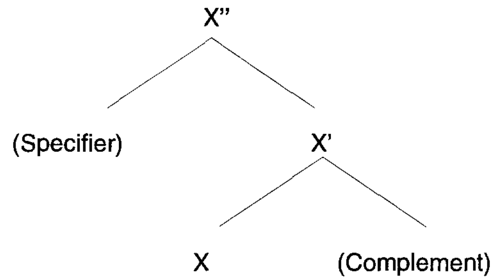

Figure A1.1. The structure of an X′ projection

- the adverb *independently* with the adverbial phrase *quite independently of me*;
- the preposition *out* with the prepositional phrase *right out of the window*;
- the verb *touches* with the verb phrase *Bill touches the pictures.*

In each case, these phrases contain a *head* (the word around which the phrase is built, i.e. *student, good, independently, out,* or *touches*), a *specifier* (a word or phrase which, in English, precedes the head and annotates its meaning, i.e. *a, very, quite, right,* or *Bill*), and a *complement* (a word or phrase which, in English, follows the head and for which the head subcategorizes, i.e. *of history, at chess, of me, of the window, the pictures*). In X′ theory, a phrase is seen as being *projected* from a head, thus the head is at the bottom of the projection—the X^0 level. Complements are attached at the X′ level, and specifiers are attached at the X″ level (often called the *maximal projection* of X).

Furthermore, phrases can also contain *adjuncts*, components which modify the meaning of the head, but which are less tightly bound to the head than are complements, e.g. compare *Bill touches the pictures gingerly* with **Bill touches gingerly the pictures*,[10] or compare *a student of history with long hair* with **a student with long hair of history.* In each of these cases, the phrases *with long hair* and *gingerly* are adjuncts. Adjuncts can attach to phrases at the X′ or the X″ level, typically by replicating the node at that level. For example, in the sentence *I consider John highly intelligent,* the phrase *John highly intelligent* is an adjectival small-clause, as shown in figure A1.2. Note how the X″ node, in this case Adj″, has been replicated in order to accommodate the adjunct, *John.*

C-command (an abbreviation of *constituent-command*) is a structural relation between nodes of an X′ tree. This relation is used in the definitions of many of the principles of generative grammar. Informally, a tree node c-commands its siblings, and the descendants of its siblings. Formally:

> A node α c-commands a node β iff α does not dominate β and the first branching node that dominates α also dominates β.

The c-command relation is then used in the definition of *government.* Government is a relation between heads and their complements, or between subjects and predicates:

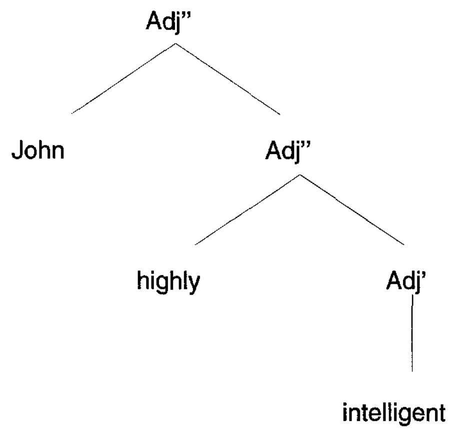

Figure A1.2. Adjunction in X′ theory

A node α governs a node β iff: α c-commands β; α is an X^0; and if δ is a maximal projection and δ dominates β then δ dominates α.

This definition simply says that in order for a node α to govern some other node β, α must be a lexical item (an X^0), it must c-command β and that α and β must be dominated by the same set of maximal projections. For example, in the sentences just cited, the noun *student* governs the prepositional phrase *of history* and the verb *touches* governs the noun phrase *the pictures*.

θ (thematic) roles are semantic roles which heads assign to their arguments when they are in a government relation. Examples of θ-roles are: *Patient, Agent, Goal, Source, Location*, etc. For example, the verb *took* has a θ-role for an agent (the thing that does the taking), a patient (the thing that was taken), and optionally a source (the place from which the patient was taken). Consider the sentence *John took cookies from orphans*. In this case, *John* is the agent, *cookies* is the patient and *from orphans* is the source.

The θ-criterion is a restriction on the θ-roles which a head can assign, and which an argument can receive:

Each head must assign one and only one θ-role to each of its arguments.
Each argument must receive one and only one θ-role.

Consider the effects of this criterion on the verb *took*. The sentences **John took cookies under orphans* is ungrammatical because the prepositional phrase *under orphans* can be a location, but it cannot be a source. Thus the head *took* can not assign all of its θ-roles and so the θ-criterion is not satisfied. Because of this the sentence is ungrammatical.

In GB theory, several levels of representation have been posited to account for the observable facts of language. Two of these levels are D (deep) structure and S (surface) structure. D-structure corresponds to the level in which pure lexical relationships are manifested. For example, in the sentence *What did Batman put in his utility belt?*, the

D-structure would be *Batman put what in his utility belt.* This structure corresponds to the observation that the verb *put* has three obligatory lexical arguments: an agent, a patient and a location. In D-structure, these arguments are represented in their canonical positions. In S-structure, however, words such as *What* are moved to their surface positions. The D-structure and S-structure positions of moved words are related, in GB theory, by the rule of Move-α. This rule simply states: Move any constituent anywhere in the X′ structure. While this sounds far too unconstrained to be useful, illegal representations which result from the application of Move-α will be ruled out by other principles, such as case marking, the θ-criterion, and so forth.

Case marking is a morphological marking on a noun which signals that noun's function in the sentence. For example, the difference between the pronoun *I* in *I see John* and the pronoun *me* in *John sees me* is that the former has nominative case and the latter has accusative case. Case, in, generative grammar, is considered to be an abstract property which is assigned by prepositions, verbs and tense (or, in the terminology used in this paper, an Agr node) to nouns, under the government relation. There is, furthermore, a restriction on sentences, called the *case filter* which states that all nouns must have case.

The notion of *agreement* as used in this paper has been informally defined as the morphological changes which can potentially occur when two syntactic structures are related in a sentence, and the phonological changes which audibly signal this relationship.

Given the terms defined in this appendix agreement can now be defined more precisely as: the morpho-phonological changes that occur on words which are in a government relationship with each other, in which one of the words either specifies or subcategorizes for the other. These morpho-phonological changes are the audible (and occasionally inaudible) analogues of meaning changes in the words—for example, number, gender, person, etc. In addition, these meaning changes can be expressed as word order in some languages with extremely simple morphology (such as Chinese). It should be noted that this definition of agreement also encompasses case assignment (Kazman, 1991c).

Appendix A2

Babel word format

The notation used to describe the input data to *Babel* is as follows: POS is *part of speech* and is one of Det(erminer), Noun, Verb, Adjective, Adverb, Neg(ator), etc.; ARGS is a set of *arguments* that the given word takes, each of which is, in turn, a set; IFEATS is a set of the *intrinsic features* of a word, such as tense or number; EFEATS is a set of the *extrinsic features*, such as the word's function within a sentence (its ROLE). The features within these sets are expressed as attribute/value pairs, taking the form: Attribute=Value.

Turning to the particular attributes used to represent the input: TH refers to *thematic* or *θ-role*, which is one of AGT (agent) or PAT (patient), GOAL, THM (theme), LOC (location), etc. PER is the *person* attribute, which may have the value 1, 2 or 3. NUM is the *number* attribute which may take the values SG (singular) or PL (plural). DEF is a boolean *definiteness* attribute. ROLE refers to the word's role in the sentence, e.g. SUB (subject), OBJ (object), 2OBJ (indirect object)). TNS, the *tense* attribute, could have the values: NPAST (non-past), INF (infinitive), PRES (present), PAST, or FUT (future). The GND, or *gender* attribute, could take the values: M (male), F (female), N (neuter), NV (non-virile), V (virile). These last two values are used in Polish plural nouns. Virile refers to men and boys. Non-virile refers to everything else. MOOD can have two values, IND (indicative) and INT (interrogative). Finally ASP, the *aspect* attribute can take the values: IMP (imperfective, also known as durative), PFV (perfective, also known as stative), or PRG (progressive).

These attributes were found to be sufficient to represent the English and Polish test data which *Babel* used—this set is by no means complete enough to describe all languages.

Notes

1. See (Langley, 1987) for a good overview of the research in this field.
2. The following terms are not those used in (Pinker, 1979), but the concepts are the same.
3. Even if children do make use of some negative evidence, there has been ample research showing that it is not a major component of their acquisition process (Newport, 1977), (Brown, 1980), (Wexler, 1980). Furthermore, a theory which works without negative evidence will always succeed with it, and so this assumption can never harm a language acquisition model.
4. *Babel* has been implemented using the Government-Binding (GB) (Chomsky, 1981) theory of syntax. For a brief explanation of the central concepts of GB theory, see appendix A1.
5. See (Abney, 1987) for an overview of the theoretical foundations which support this view of language.
6. Affixes in this paper will be represented in their most common orthographic form, (e.g. *-ed* for the past tense affix), rather than in their phonological form, for readability. *Babel* deals only with phonological forms however. In addition, an affix such as *-ed*, when used in the text, is meant to symbolize all of the phonological variants which that affix could take, in this case /d/, /t/, /Id/.
7. Semantically related words are, for instance, *walk* and *walks*, but not *walk* and *stroll*. This relationship is coded in the input, not induced by the system.
8. I am borrowing the notion of *deductive acceptability* from Horning (Horning, 1969). Intuitively, this notion means that a hypothesis which is deductively falsifiable (given the child's current state of linguistic knowledge) should never be inferred by his learning procedure, based upon its current state. This means that rules will only apply in appropriate settings, when their phonological and categorial preconditions are met. For example, the plural *-s* morpheme in English agrees in voicing with the preceding phoneme, and so it would be deductively *un*acceptable to have the child hypothesize a voiced plural affix *-z* attached to an unvoiced stem, like *book*. Such a rule application would be deductively falsifiable given the child's current state of knowledge, and thus ruled out. The notion of deductive acceptability not only greatly reduces the size of the inference problem for the child, but also accords with observations of children learning the morphology of their language—they freely apply productive affixation rules, even to the point of overgeneralizing the use of productive affixes, but *only* under the proper phonological conditions.
9. When morphemes became productive at the same time, they were given the same rank by Brown and the de Villiers. This is why, in the de Villiers' study, the present progressive, "on" and "in" all have the same rank of 2.
10. Preceding a sentence with a $*$ is a linguistic convention indicating ungrammaticality.

References

Abney, Steven P. *The English Noun Phrase in its Sentential Aspect.* PhD thesis, Massachusetts Institute of Technology, 1987.

Anderson, John. Language acquisition by computer and child. Technical Report Human Performance Center # 55, University of Michigan, 1974.

Anderson, John. Computer simulation of a language acquisition system: A first report. In R. Solso, editor, *Information Processing and Cognition: The Loyola Symposium.* Lawrence Erlbaum, Washington, 1975.

Baker, Mark and Hale, Kenneth. Relativized minimality and pronoun incorporation. *Linguistic Inquiry*, 21(2):289–297, 1990.

Berman, Ruth A. The acquisition of Hebrew. In Dan I. Slobin, editor, *The Cross-linguistic Study of Language Acquisition, Volume 1: The Data*, pages 255–371. Lawrence Erlbaum, Hillsdale, New Jersey, 1985.

Berwick, Robert C. *The Acquisition of Syntactic Knowledge.* MIT Press, Cambridge, Mass., 1985.

Brown, Roger and Hanlon, Camille. Derivational complexity and order of acquisition in child speech. In J. R. Hayes, editor, *Cognition and the development of language.* Wiley, New York, 1970.

Brown, Roger. *A First Language: The Early Stages.* Harvard University Press, Cambridge, Mass., 1973.

Caplan, David and Hildenbrandt, Nancy. *Disorders of Syntactic Comprehension.* MIT Press, Cambridge, Mass., 1988.

Chomsky, Noam. *Lectures on Government and Binding.* Foris, Dordrecht, Holland, 1981.

Chomsky, Noam. Some notes on economy of derivation and representation. In Itziar Laka and Anoop Mahajan, editors, *MIT Working Papers in Linguistics 10: Functional Heads and Clause Structure.* Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 1989.

Chomsky, Noam and Halle, Morris. *The Sound Pattern of English.* Harper and Row, New York, 1968.

Clahsen, Harald. Critical phases of grammar development. A study of the acquisition of negation in children and adults. In Peter Jordens and Josien Lalleman, editors, *Language Development*, pages 123–148. Foris, Dordrecht, Holland, 1988.

Clark, Eve. The Acquisition of Romance, with Special Reference to French. In Dan I. Slobin, editor, *The Cross-linguistic Study of Language Acquisition, Volume 1: The Data*, pages 687–782. Lawrence Erlbaum, Hillsdale, New Jersey, 1985.

Davis, Henry and LeBlanc, David. A model of the development of phrase structure. Working Notes: AAAI Spring Symposium Series, Workshop on Machine Learning of Natural Language and Ontology, 1991.

de Villiers, Jill G. and de Villiers, Peter A.. The Acquisition of English. In Dan I. Slobin, editor, *The Cross-linguistic Study of Language Acquisition, Volume 1: The Data*, pages 27–139. Lawrence Erlbaum, Hillsdale, New Jersey, 1985.

Emmorey, Karen D. *Morphological Structure and Parsing in the Lexicon.* PhD thesis, University of California, Los Angeles, 1987.

Emonds, J. E. *A Unified Theory of Syntactic Categories.* Foris, Dordrecht, Holland, 1985.

Ervin, Susan. Imitation in children's language. In E. H. Lenneberg, editor, *New Directions in the Study of Language.* MIT Press, Cambridge, Mass., 1964.

Fong, Sandiway. The computational implementation of principle-based parsers. In C. Tenny, editor, *The MIT Parsing Volume, 1988-1989.* MIT Center for Cognitive Science, Cambridge, Mass., 1989.

Fukui, Naoki and Speas, Margaret. Specifiers and projections. ms. MIT, 1987.

Gibson, Edward. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown.* PhD thesis, Carnegie Mellon University, 1991.

Grimshaw, Jane. Form, function, and the language acquisition device. In C. L. Baker and John J. McCarthy, editors, *The Logical Problem of Language Acquisition*, pages 165–182. MIT Press, Cambridge, Mass., 1981.

Horning, James. A study of grammatical inference. Stanford University Computer Science Department Technical Report No. CS139, 1969.

Kazman, Rick. The Genesis of Functional Categories. Presented to the 15th Boston University Conference on Language Development, 1990.

Kazman, Rick. On building a model of grammar from information in the lexicon. Working Notes: AAAI Spring Symposium Series, Workshop on Machine Learning of Natural Language and Ontology, 1991.

Kazman, Rick. Babel: A psychologically plausible cross-linguistic model of lexical and syntactic acquisition. In L. A. Birnbaum and G. C. Collins, editors, *Proceedings of the Eighth International Workshop (ML91)*, pages 75–79, San Mateo, CA, 1991. Morgan Kaufmann.

Kazman, Rick. Why do children say "Me do it"? In K. J. Hammond and D. Gentner, editors, *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, pages 455–460, Hillsdale, New Jersey, 1991. Lawrence Erlbaum.

Kazman, Rick. *The Induction of the Lexicon and the Early Stages of Grammar.* PhD thesis, Carnegie Mellon University, 1991.

Lachter, Joel and Bever, Thomas G. Language and connectionism. In Stephen Pinker and Jacques Mehler, editors, *Connections and Symbols*, pages 193–247. MIT Press, Cambridge, Mass., 1988.

Langley, Pat and Carbonell, Jaime G. Language acquisition and machine learning. In Brian MacWhinney, editor, *Mechanisms of Language Acquisition*, pages 115–155. Lawrence Erlbaum, Hillsdale, New Jersey, 1987.

MacWhinney, Brian and Snow, Catherine. The child language data exchange system. *Journal of Computational Linguistics*, 12:271–296, 1985.

Marcus, Gary F., Ullman, Michael, Pinker, Steven, Hollander, Michelle, Rosen, T. John, and Xu, Fei. Overregularization. Technical Report Center for Cognitive Science Occasional Paper #41, Massachusetts Institute of Technology, 1990.

Newell, E., Gleitman, H., and Gleitman, L. Mother, please, I'd rather do it myself: some effects and non-effects of maternal speech style. In C. Snow and C. Ferguson, editors, *Talking to Children: Language Input and Acquisition*, pages 109–150. Cambridge University Press, New York, 1977.

Pinker, Stephen. Formal models of language learning. *Cognition*, (1):217–283, 1979.

Pinker, Stephen. *Language Learnability and Language Development.* Harvard University Press, Cambridge, Mass., 1984.

Pinker, Stephen and Prince, Alan. Language and connectionism. In Stephen Pinker and Jacques Mehler, editors, *Connections and Symbols*, pages 73–193. MIT Press, Cambridge, Mass., 1988.

Radford, Andrew. *Transformational Grammar: A First Course.* Cambridge University Press, Cambridge, UK, 1988.

Rumelhart, David and McClelland, James. On learning the past tenses of English verbs. In J. L. McClelland and D. E. Rumelhart, editors, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Bradford Books/MIT Press, Cambridge, Mass., 1986.

Scarborough, D. L. , Cortese, C., and Scarborough, H. S. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3:1–17, 1977.

Selfridge, Mallory. A computer model of child language learning. In *Proceedings of the 1st annual conference of the American Association for Artificial Intelligence*, 1980.

Selfridge, Mallory. Why do children say "goed"? A computer model of child generation. In *Proceedings of the 3rd Annual Meeting of the Cognitive Science Society*, 1981.

Slobin, Dan I. On the nature of talk to children. In E. H. Lenneberg and E. Lenneberg, editors, *Foundations of Language Development.* Academic Press, New York, 1975.

Smoczyńska, Magdalena. The Acquisition of Polish. In Dan I. Slobin, editor, *The Cross-linguistic Study of Language Acquisition, Volume 1: The Data*, pages 595–686. Lawrence Erlbaum, Hillsdale, New Jersey, 1985.

Suppes, Patrick. The semantics of children's language. *American Psychologist*, pages 103–114, February 1973.

Weist, Richard, Wysocka, Hanna, Witkowska-Stadnik, Katarzyna, Buczowska, Ewa, and Koneiczna, Emilia. The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language*, 11:347–374, 1984.

Wexler, Kenneth and Culicover, Peter. *Formal Principles of Language Acquisition.* MIT Press, Cambridge, Mass., 1980.

Received March 25, 1992

Accepted January 25, 1994