# Item-based Constructions and the Logical Problem

**Brian MacWhinney**
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
macw@cmu.edu

## Abstract

The logical problem of language is grounded on arguments from poverty of positive evidence and arguments from poverty of negative evidence. Careful analysis of child language corpora shows that, if one assumes that children learn through item-based constructions, there is an abundance of positive evidence. Arguments regarding the poverty of negative evidence can also be addressed by the mechanism of conservative item-based learning. When conservativism is abandoned, children can rely on competition, cue construction, monitoring and probabilistic identification to derive information from positive data to recover from overgeneralization.

## 1. The Logical Problem

Chomsky (1957, 1980) has argued that the child's acquisition of grammar is 'hopelessly underdetermined by the fragmentary evidence available.' He attributed this indeterminacy to two major sources. The first is the degenerate nature of the input. According to Chomsky, the sentences heard by the child are so full of retracing, error, and incompletion that they provide no clear indication of the possible sentences of the language. Coupled with this problem of input degeneracy is the problem of unavailability of negative evidence. According to this view, children have a hard time knowing which forms of their language are acceptable and which are unacceptable, because parents fail to provide consistent evidence regarding the ungrammaticality of unacceptable sentences. Worse

still, when such evidence is provided, children appear to ignore it.

Chomsky's (1957) views about the degeneracy of the input did not stand up well to the test of time. As Newport, Gleitman & Gleitman (1977) reported, 'the speech of mothers to children is unswervingly well-formed.' More recently, Sagae, Lavie & MacWhinney (2004) examined several of the corpora in the CHILDES database and found that adult input to children can be parsed with an accuracy level parallel to that for corpora such as the Wall Street Journal database.

This evidence for well formedness of the input did not lead to the collapse of the 'argument from poverty of stimulus' (APS). However, it did place increased weight on the remaining claims regarding the absence of relevant evidence. The overall claim is that, given the absence of appropriate positive and negative evidence, no child can acquire language without guidance from a rich set of species-specific innate hypotheses. Some refer to the argument from poverty of stimulus as the 'logical problem of language acquisition (Baker, 1979), while others have called it 'Plato's Problem,' 'Chomsky's Problem,' 'Gold's Problem,' or 'Baker's Paradox.'

## 2. Absence of Negative Evidence

In the 1970s, generativist analyses of learnability (Wexler & Hamburger, 1973) relied primarily on an analysis presented by Gold (1967). Gold's analysis contrasted two different language-learning situations: text presentation and informant presentation. With informant presentation, the language learner can receive feedback from an infallible informant regarding the grammaticality of every candidate sentence. This corrective feedback is called 'negative evidence' and it only requires that

ungrammatical strings be clearly identified as unacceptable. Whenever the learner formulates an overly general guess about some particular linguistic structure, the informant will label the resulting structure as ungrammatical and the learner will use this information to restrict the developing grammar. Based on initial empirical results reported by Brown & Hanlon (1970), Gold argued that negative evidence is not available to the child and that language learning cannot be based on informant presentation.

Marcus (1993) has argued that the feedback that parents provide does not discriminate consistently between grammatical and ungrammatical constructions. As a result, children cannot rely on simple, overt negative evidence for recovery from overgeneralization. Although I will argue that parents provide positive evidence in a form that solves the logical problem (Bohannon *et al.*, 1990), I agree with the observation that this evidence does not constitute overt grammatical correction of the type envisioned by Gold.

## 3. Absence of Positive Evidence

Beginning about 1980, generative analyses of learnability began to shift away from an emphasis on the unavailability of negative evidence to arguments based on the unavailability of positive evidence. This conceptual shift led to a relative decline in attention to recovery from overgeneralization and an increase in attention to reported cases of error-free learning. For example, Chomsky's (1980) statement of the logical problem relies on the notion of error-free learning without positive evidence. The argumentation here is that, if a structure is never encountered in the input, correct use of this structure would have to indicate innate knowledge.

Researchers have claimed that the child produces error-free learning without receiving positive evidence for structures such as: structural dependency, c-command, the binding conditions, subjacency, negative polarity items, that-trace deletion, nominal compound formation, control, auxiliary phrase ordering, and the empty category principle. In each of these cases, it is necessary to assume that the underlying universal is a part of the non-parameterized core of universal grammar. If the dimension involved were parameterized, there would be a need for some form of very early pa-

rameter setting (Wexler, 1998), which could itself introduce some error. Thus, we would expect error-free learning to occur primarily for those aspects of the grammar that are completely universal and not parameterized. Parameterized features, such as subject pro-drop, could still be guided by universal grammar. However, their learning would not necessarily be error-free.

### 3.1. Structural dependency

The paradigm case of error-free learning is the child's obedience to the Structural Dependency condition, as outlined by Chomsky in his formal discussion with Jean Piaget (Piattelli-Palmarini, 1980). Chomsky notes that children learn early on to move the auxiliary to initial position in questions, such as 'Is the man coming?' One formulation of this rule is that it stipulates the movement of the first auxiliary to initial position. This formulation would be based on surface order, rather than structural relations. However, if children want to question the proposition given in (1), they will never produce a movement such as (2). Instead, they will always produce (3).

1. The man who is running is coming.
2. Is the man who __ running is coming?
3. Is the man who is running __ coming?'

In order to produce (3), children must be basing the movement on structure, rather than surface order. Thus, according to Chomsky, they must be innately guided to formulate rules in terms of structure.

In the theory of barriers (Chomsky, 1986), the repositioning of the auxiliary in the tree and then in surface structure involves a movement of INFL to COMP that is subject to the head movement constraint. In (2) the auxiliary would need to move around the N' of 'man' and the CP and COMP of the relative clause, but this movement would be blocked by the head movement constraint (HMC). No such barriers exist in the main clause. In addition, if the auxiliary moves as in (2), it leaves a gap that will violate the empty category principle (ECP). Chomsky's discussion with Piaget does not rely on these details. Chomsky simply argues that the child has to realize that phrasal structure is somehow involved in this process and that one cannot formulate the rule of auxiliary movement as 'move the first auxiliary to the front.'

Chomsky claims that, 'A person might go through much or all of his life without ever having been exposed to relevant evidence, but he will nevertheless unerringly employ the structure-dependent generalization, on the first relevant occasion.' A more general statement of this type provided by Hornstein & Lightfoot (1981) who claim that, 'People attain knowledge of the structure of their language for which no evidence is available in the data to which they are exposed as children.'

In order to evaluate these claims empirically, we need to know when children first produce such sentences and whether they have been exposed to relevant examples in the input prior to this time. In searching for instances of relevant input as well as first uses, we should include two types of sentences. First, we want to include sentences such as (3) in which the moved verb was a copula in the relative clause, as well as sentences with auxiliaries in both positions, such as 'Will the boy who is wearing a Yankee's cap step forward?' The auxiliaries do not have to be lexically identical, since Chomsky's argument from poverty of stimulus would also apply to a child who was learning the movement rule on the basis of lexical class, as opposed to surface lexical form.

Examining the TreeBank structures for the Wall Street Journal in the Penn TreeBank, Pullum & Scholz (Pullum & Scholz, 2002) estimate that adult corpora contain up to 1% of such sentences. However, the presence of such structures in formal written English says little about their presence in the input to the language-learning child. A search by Lewis & Elman (2001) of the input to English-speaking children in the CHILDES database (MacWhinney, 2000) turned up only one case of this structure out of approximately 3 million utterances. Since CHILDES includes good sampling of target children up to age 5;0, we can safely say that positive evidence for this particular structure is seldom encountered in the language addressed to children younger than 5;0.

Because children do not produce sentences of this type themselves, it is difficult to use production data to demonstrate the presence of the constraint. Crain & Nakayama (1987) attempted to get around this problem by eliciting these forms from children directly. They asked children (3;2 to 5;11) to, 'Ask Jabba if the boy who is watching Mickey is happy.' Children responded with a variety of structures, none of which involved the movement of the auxiliary from the relative clause. Unfortunately, this elicitation procedure encourages children to treat the relative clause ('the boy who is watching Mickey') as an imitated chunk. Despite the serious methodological limitation in this particular study, it seems reasonable to believe that four-year-old children are beginning to behave in accordance with the Structural Dependency condition for sentences like (2) and (3). But does this mean that they reach this point without learning?

There is another type of sentence that provides equally useful positive evidence regarding auxiliary movement. These are wh-questions with embedded relative clauses. It turns out that there are hundreds of input sentences of this type in the CHILDES corpus. Most of these have the form of (4), but some take the form of (5).

4. Where is the dog that you like?
5. Which is the dog that is clawing at the door?

In (5) the child receives clear information demonstrating that moved auxiliaries derive from the main clause and not the relative clause. Using evidence of the type provided in (4), the child simply learns that moved auxiliaries and the wh-words that accompany them are arguments of the verb of the main clause. Sentences like (4) and (5) are highly frequent in the input to children and both types instruct the child in the same correct generalization.

Based on evidence from the main clause, the child could formulate the rule as a placement after the wh-word of the auxiliary that is conceptually related to the verb being questioned. In other words, it is an attachment to the wh-word of an argument of the main verb. This is a complex application of the process of item-based construction generation proposed in MacWhinney (1975, 1982). This formulation does not rely on barriers, ECP, HCP, INFL, COMP, or movement. It does rely on the notion of argument structure, but only as it emerges from the application of item-based constructions. Given this formulation, a few simple yes–no questions would be enough to demonstrate the pattern. When children hear 'is the baby happy' they can learn that the initial copula auxiliary 'is' takes a subject argument in the next slot and a predicate argument in the following slot. They will learn similar frames for each of the other fronted auxiliaries. When they then encounter sen-

tences such as (11) and (12), they will further elaborate the item-based auxiliary frames to allow for positioning of the initial wh-words and for attachment of the auxiliaries to these wh-words.

One might argue that this learning scenario amounts to a restatement of Chomsky's claim, since it requires the child to pay attention to relational patterns, rather than serial order as calculated from the beginning of the sentence. However, if the substance of Chomsky's claim is that children learn to fill argument slots with compound constituents, then his analysis seems indistinguishable from that of MacWhinney (1975; 1987a).

## 3.2 Auxiliary phrases

Kimball (1973) presented perhaps the first example of a learnability problem based on poverty of positive evidence. He noted that children are exposed to scores of sentences with zero, one, or two auxiliaries as in (6)–(13). However, his searches of a million sentences in early machine-readable corpora located not a single instance of a structure such as (13).

6.  It rains.
7.  It may rain.
8.  It may have rained.
9.  It may be raining.
10. It has rained.
11. It has been raining.
12. It is raining.
13. It may have been raining.

Kimball argued that, despite the absence of positive data for (13), children are still able to infer its grammaticality from the data in (6) to (12). He took this as evidence that children have innate knowledge of structural compositionality. The empirical problem with Kimball's analysis is that sentences like (13) are not nearly as rare as his corpus analysis suggests. My search of the CHILDES database for the string 'might have been' located 27 instances in roughly 3 million sentences. In addition there were 24 cases of 'could have been', 15 cases of 'should have been', and 70 cases of 'would have been.' Thus, there seems to be little shortage of positive evidence for the direct learning of this pattern. Perhaps Kimball's findings to the contrary arose from focusing exclusively on 'may', since a search for 'may have been' turned up only 5 cases.

## 3.3 The complex-NP constraint

The complex-NP constraint blocks movement of a noun from a relative clause, as in (14) and (15).

14. *Who did John believe the man that kissed __ arrived
15. Who did John believe __ kissed his buddy?

This same constraint also blocks movement from prepositional phrases and other complex NPs, as in (16) – (18):

16. *Who did pictures of ___ surprise you?
17. *What did you see a happy ___ ?
18. *What did you stand between the wall and ___ ?

The constraint in (18) has also been treated as the coordinated-NP constraint in some accounts. Although it appears that most children obey these constraints, there are some exceptions. Wilson & Peters (1988) list these violations of the complex NP constraint from Wilson's son Seth between the ages of 3;0 and 5;0.

19. What am I cooking on a hot __ ? (stove)
20. What are we gonna look for some __ ? (houses)
21. What is this a funny __ , Dad?
22. What are we gonna push number __ ? (9)
23. Where did you pin this on my __ ? (robe)
24. What are you shaking all the __ ? (batter and milk)
25. What is this medicine for my __ ? (cold)

These seven violations all involve separation of a noun from its modifiers. Two other examples, illustrate violation of the complex-NP constraint in other environments:

26. What did I get lost at the __ , Dad?
27. What are we gonna go at Auntie and __ ?

Here, the prohibited raising involves prepositional phrases and a conjoined noun phrase. Violations of the latter type are particularly rare, but still do occur occasionally.

One might object that a theory of universal grammar should not be rejected on the basis of a few violations from a single child. However, other observers have reported similar errors. In the recordings from my sons Ross and Mark, I observed a few such violations. One occurred when my son Mark (at 5;4.4) said, 'Dad, next time when it's Indian Guides and my birthday, what do you think a picture of ___ should be on my cake?' Catherine Snow reports that at age 10;10, her son Nathaniel said, 'I have a fever, but I don't want to

said, 'I have a fever, but I don't want to be taken a temperature of.'

Most researchers would agree that violations of the complex-NP constraint are rare, but certainly not nonexistent. At the same time, the structures or meanings that might trigger these violations are also very rare, as is the input that would tell the child how to handle these structures. Given this, it seems to me that these patterns cannot reasonably be described as cases of error-free learning. Instead, we should treat them as instances of 'low-error constructions.' In this regard, they resemble errors such as stative progressives ('I am knowing') and double-object violations ('He recommended the library the book'). As soon as we shift from error-free learning to low-error learning, we need to apply a very different form of analysis, since we now have to explain how children recover from making these overgeneralization errors, once they have produced them. This then induces us to again focus on the availability of negative evidence.

Of course, we could assume that the violation of the complex-NP constraint was a transient performance error and that, once the relevant performance factors are eliminated, the constraints of UG operate to block further wh-raising from complex noun phrases. But the important point here is that we now need to consider specific mechanisms for allowing for recovery from overgeneralization, even for what have been offered as the clearest cases of the application of universal constraints.

## 3.4 Binding conditions

Binding theory (Chomsky, 1981) offers three proposed universal conditions on the binding of pronouns and reflexives to referents. Sentence (28) illustrates two of the constraints. In (28), 'he' cannot be coreferential with 'Bill' because 'Bill' does not c-command the pronoun. At the same time, 'himself' must be coreferential with 'Bill' because it is a clausemate and does c-command 'Bill.'

28.   He said that Bill hurt himself.

When attempting to relate the logical problem to the study of the binding constraints, it is important to remember that the sentences produced or interpreted are fully grammatical. However, the interpretation in which the pronoun is coreferential with the full NP is disallowed by the binding principles. This means that, to study the imposition of the constraints, researchers must rely on comprehension studies, often with very young children.

It is well known that children often fail to apply these principles, even in carefully controlled experiments (O'Grady, 1997). Various accounts have been offered to reconcile these facts with the supposed universality of the constraint. However, one possibility that has seldom been explored is the idea that the binding conditions are learned on the basis of positive data. To illustrate the role that learning can play in this area, consider a study of long-distance movement of adjuncts by De Villiers, Roeper & Vainikka (De Villiers *et al.*, 1990). Children were divided into two age groups: 3;7 to 5;0 and 5;1 to 6;11. They were given sentences such as:

29.   When did the boy say he hurt himself?
30.   When did the boy say how he hurt himself?
31.   Who did the boy ask what to throw?

For (29), 44% of the children gave long distance interpretations, associating 'when' with 'hurt himself', rather than 'say.' For (30), with a medial wh-phrase blocking a long-distance interpretation, only 6% gave long-distance responses. This shows that children were sensitive to the conditions on traces, in accord with P&P (Chomsky & Lasnik, 1993) theory. However, the fact that sensitivity to this contrast increases markedly across the two age groups indicates that children are learning this pattern. In the youngest group, children had trouble even understanding sentences with medial arguments like (31). The fact that this ability improves over time again points to learning of the possible interpretations of these structures.

Children can learn to interpret these sentences correctly by applying conservative learning principles that rely on positive data. First, they learn short-distance interpretations that attach the wh-word to the main clause. Then, when they hear sentences with medial "how" they add the additional possibility of the long-distance interpretation. However, they do this in a conservative item-based manner, limiting the new interpretation to sentences like (30) with medial "how."

P&P theory can also provide an account of this development in terms of the setting of parameters. First, children must realize that their language allows movement, unlike Chinese. Next they must decide whether the movement can be local, as in German, or both local and distant as in English.

Finally, they must decide whether the movement is indexed by pronouns, traces, or both. However, once a parameter-setting account is detailed in a way that requires careful attention to complex cue patterns over time (Buttery, 2004; Sakas & Fodor, 2001), it can be difficult to distinguish it from a learning account. Using positive evidence, children can first learn that some movement can occur. Next, they can learn to move locally and finally they can acquire the cues to linking the moved argument to its original argument position, one by one.

## 3.5 Learnability or learning?

What have we learned from our examination of these four examples? First, we have seen that the application of universal constraints is not error-free. This is particularly true in the case of the binding conditions. Because the binding conditions involve parameter setting, it is perhaps not surprising that we see errors in this domain. However, we also find errors in the application of the non-parameterized constraint against raising from complex noun phrases. Only in the case of the structural dependency condition do we find no errors. However, for that structure there is also no usage at all by either parents or children, unless we consider attachment of auxiliaries to wh-words, which is quite frequent. It is possible that error-free learning exists in various other corners of syntactic, semantic, or lexical learning. But there is no evidence that error-free learning occurs in association with an absence of positive evidence. This is the crucial association that has been claimed in the literature and it is the one that we have shown to be false.

Second, for each of the four learnability problems we examined, we have seen that there are effective learning methods based on available positive evidence. This learning involves mechanisms of conservative, item-based learning followed by later generalization.

## 4. Multiple Solutions

Having now briefly surveyed the role of the logical problem in generative theory, we turn next to a consideration of seven factors that, operating together, allow the child to solve the logical problem. Of these seven factors, the first two are simply formal considerations that help us understand the scope of the problem. The last five are processes that can actually guide the child during acquisition.

## 4.1 Limiting the class of grammars

The first solution to the logical problem addresses the Gold analysis directly by showing how language can be generated from finite-state grammars (Reich, 1969). For example, Hausser (1999) has developed an efficient parser for left-associative grammars. He has shown that left-associative grammar can be expressed as a finite automaton that orders words in terms of part-of-speech categories. Because we know that finite automata can be identified from positive evidence (Hopcroft & Ullman, 1979), this means that children should be able to learn left-associative grammars directly without triggering a logical problem. Given the fact that these grammars can parse sentences in a time-linear and psycholinguistically plausible fashion, they would seem to be excellent candidates for further exploration by child language researchers.

A formal solution to the logical problem also arises in the context of the theory of categorical grammar. Kanazawa (1998) shows that a particular class of categorial grammars known as the k-valued grammars can be learned on positive data. Moreover, he shows that most of the customary versions of categorial grammar discussed in the linguistic literature can be included in this k-valued class. Shinohara (1994) and Jain, Osherson, Royer & Sharma (1999) examine still further classes of complex non-finite languages that can be learned on the basis of positive data alone. These attempts to recharacterize the nature of human language by revised formal analysis all stand as useful approaches to the logical problem. By characterizing the target language in a way that makes it learnable by children, linguists help bridge the gap between linguistic theory and child language studies.

## 4.2 Revised end-state criterion

The second solution to the logical problem involves resetting our notion of what it means to acquire an end-state grammar. Horning (1969) showed that, if the language identification is allowed to involve a stochastic probability of identification, rather than an absolute guarantee of no further error ever, then language can be identified on positive evidence alone. It is surprising that this

solution has not received more attention, since this analysis undercuts the core logic of the logical problem, as it applies to the learning of all rule systems up to the level of context-sensitive grammars. If learning were deterministic, children would go through a series of attempts to hypothesize the 'correct' grammar for the language. Once they hit on the correct identification, they would then never abandon this end-state grammar. The fact that adults make speech errors and differ in their judgments regarding at least some syntactic structures suggests that this criterion is too strong and that the view of grammar as stochastic is more realistic.

## 4.3 Conservative Item-based Learning

The third solution to the logical problem emphasizes the conservative nature of children's language learning. The most direct way for a language learner to solve Gold's problem is to avoid formulating overly general grammars in the first place. If the child never overgeneralizes, there is no problem of recovery from overgeneralization and no need for negative evidence or corrective feedback. Taking this basic idea one step further, let us imagine that grammars are ordered strictly in terms of their relative generative power. If this is true, then the forms generated by a grammar are a subset of the next slightly larger grammar. This is known as the Subset Principle. If the child always chooses the least powerful grammar that is consistent with the input data, then the problem of the unavailability of negative evidence disappears and learning can be based simply on positive evidence.

The Subset Principle has often been used to argue for abstract relations between grammars. For example, Fodor & Crain (1987) argue that the child learns the periphrastic dative ('give the book to John') for each new verb and only assumes that the double object construction ('give John the book') can be applied if it is attested in the input. In this particular case, the grammar with only the periphrastic is ordered as a subset of the grammar with both constructions. This follows from the principles for expansion of curly braces in GPSG.

Conservatism can control acquisition of these structures without invoking the Subset Principle. The theory of item-based acquisition (MacWhinney, 1975, 1982, 1987a; Tomasello, 2000) holds that syntactic learning is driven by the induction and combination of item-based construc-

tions. Each item-based construction specifies a set of slots for arguments. Initially, these slots encode features that are specific to the first words encountered in this slot during comprehension. For example, the item 'more' has a slot for a following argument. If the first combinations the child picks up from comprehension are 'more cookies' and 'more milk', then this slot will initially be limited to foods. However, as the child hears 'more' used in additional combinations, the semantics of the slot filler will extend to any mass noun or plural. This learning is based entirely on generalization from positive evidence.

When learning the item-based construction for 'give', children encounter sentences such as 'Bill gives John the book.' From this, they learn the double-object construction: giver + 'give' + recipient + gift. They also learn the competing item-based construction of giver + 'give' + gift + 'to' recipient. There is no need to invoke the Subset Principle to explain this learning, since item-based constructions are inherently conservative and provide their own constraints on the form of grammars. Having acquired these two basic constructions, children can them join them into a single item-based finite automaton that operates on narrowly defined lexical categories.
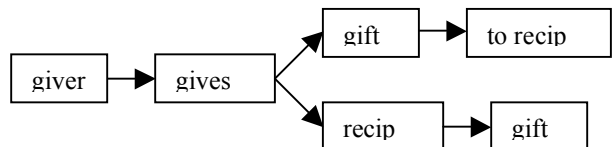


Children can learn this item-based grammar fragment on the basis of simple positive data. This example uses the formalism of a finite-state automaton to annotate the use of positive data. However, in the Competition Model and other connectionist accounts, the two verb frames compete probabilistically with the outcome of the competition being determined by further cues such as focusing or topicalization.

Item-based learning involves an ongoing process of generalization for the semantic features of the arguments. During these processes of generalization, to minimize the possibility of error, the child has to be conservative in three ways:

- The child needs to formulate each syntactic combination as an item-based construction.

- Each item-based construction needs to record the exact semantic status of each positive instance of an argument in a particular grammatical configuration (MacWhinney, 1987a).
- Attempts to use the item-based construction with new arguments must be closely guided by the semantics of previously encountered positive instances.

If the child has a good memory and applies this method cautiously, overgeneralization will be minimized and there will be no need to recover from overgeneralization.

Each item-based construction is linked to a specific lexical item. This item must be a predicate. There are no item-based constructions for nouns. Predicates can have up to three arguments. Item-based constructions for verbs can also include the verbs of embedded clauses as possible arguments. Item-based constructions for prepositions and auxiliaries include both a phrase internal head (endohead) and a head for the phrase attachment (exohead). For details on the implementation of this grammatical relations model through a parser see Sagae, MacWhinney, and Lavie (2004). In section 4.6, we will see how item-based constructions are generalized to feature-based constructions in accord with the account of MacWhinney (1987a)

Conservatism also applies to non-local movement patterns. For example, Wolfe Quintero (1992) has shown that conservatism can be used to account for L2 acquisition of the wh-movement patterns. She notes that L2 learners acquire these positive contexts for wh-movement in this order:

32. What did the little girl hit __ with the block today?
33. What did the boy play with __ behind his mother?
34. What did the boy read a story about __ this morning?

Because they are proceeding conservatively, learners never produce forms such as (35):

35. *What did the boy with ___ read a story this morning?

They never hear this structure in the input and never hypothesize a grammar that includes it. As a result, they never make overgeneralizations and never attempt wh-movement in this particular context. Data from Maratsos, Kuczaj, Fox & Chalkley (1979) show that this same analysis applies to first language learners.

## 4.4 Competition

Conservatism is a powerful mechanism for addressing the logical problem. However, children will eventually go 'beyond the information given' and produce errors (Jespersen, 1922). When the child produces errors, some mechanism must force recovery. The four processes that have been proposed by emergentist theory are: competition, cue construction, monitoring, and indirect negative evidence. Each of these processes can work to correct overgeneralization. These processes are important for addressing the version of the logical problem that emphasizes the poverty of negative evidence.

The fourth solution to the problem of poverty of negative evidence relies on the mechanism of competition. Of the four mechanisms for promoting recovery from overgeneralization, competition is the most basic, general, and powerful. Psychological theories have often made reference to the notion of competition. In the area of language acquisition, MacWhinney (1978) used competition to account for the interplay between 'rote' and 'analogy' in learning morphophonology. Competition was later generalized to all levels of linguistic processing in the Competition Model. In the 1990s, specific aspects of learning in the Competition Model were formulated through both neural network theory and the ACT-R production system.

The Competition Model views overgeneralizations as arising from two types of pressures. The first pressure is the underlying analogic force that produces the overgeneralization. The second pressure is the growth in the rote episodic auditory representation of a correct form. This representation slowly grows in strength over time, as it is repeatedly strengthened through encounters with the input data. These two forces compete for the control of production. Consider the case of '*goed' and 'went'. The overgeneralization 'goed' is supported by analogy. It competes against the weak rote form 'went,' which is supported by auditory memory. As the strength of the rote auditory form for 'went' grows, it begins to win out in the competition against the analogic form '*goed'. Finally, the error is eliminated. This is the Competition Model account for recovery from overgeneralization. The competition between two candidate forms is governed by the strength of their episodic auditory representations. In the case of the competition be-

tween '*goed' and 'went', the overgeneralized form has little episodic auditory strength, since it is heard seldom if at all in the input. Although '*goed' lacks auditory support, it has strong analogic support from the general pattern for past tense formation. In the Competition Model, analogic pressure stimulates overgeneralization and episodic auditory encoding reins it in. The analogic pressure hypothesized in this account has been described in detail in several connectionist models of morphophonological learning. The models that most closely implement the type of competition being described here are the models of MacWhinney and Leinbach (1991) for English and MacWhinney, Leinbach, Taraban & McDonald (1989) for German. In these models, there is a pressure for regularization according to the general pattern that produces forms such as '*goed' and '*ranned'. In addition, there are weaker gang effects that lead to overgeneralizations such as '*stang' for the past tense of 'sting'.

Competition implements the notion of blocking developed first by Baker (1979) and later by Pinker (1994). Blocking is more limited than competition because it requires either strict rule-ordering or all-or-none competition. The assumption that forms are competing for the same meaning is identical to the Principle of Uniqueness postulated by Pinker (1994). Competition is also the general case of the Direct Contrast noted by Saxton (1997).

Competition goes beyond the analyses offered by Baker, Pinker, and Saxton by emphasizing the fact that the child is continually internalizing adult forms in episodic memory. Recent evidence for the power of episodic memory in infant audition (Aslin *et al.*, 1999) has underscored the power of neural mechanisms for storing linguistic input and extracting patterns from this input without conscious processing. The Competition Model assumes that children are continually storing traces of the words and phrases they hear along with tags that indicate that these phrases derive directly from adult input. When the child then comes to produce a spontaneous form, these stored forms function as an 'oracle' or 'informant', providing delayed negative evidence that corresponds (because of competition or Uniqueness) to the currently generated productive form. The ultimate source of this negative evidence is the input. Children do not use this evidence when it is initially presented. It is only later when the information is retrieved in the con-

text of productive combinations that it provides negative evidence. This can only happen if it is clear that stored adult forms compete directly (Saxton, 1997) with productive child forms. The crucial claim of the Competition Model is that the same retrieval cues that trigger the formation of the overgeneralized productive form also trigger the retrieval of the internalized negative evidence. When these assumptions hold, there is a direct solution to the logical problem through the availability of internalized negative evidence.

To gain a better understanding of the range of phenomena that can be understood in terms of competition, let us look at examples from morphology. lexical semantics, and syntactic constructions.

### 4.4.1 Morphological competition

Bowerman (1987) argued that recovery from overgeneralizations such as '*unsqueeze' is particularly problematic for a Competition Model account. She holds that recovery depends on processes of semantic reorganization that lie outside the scope of competition. To make her example fully concrete, let us imagine that '*unsqueeze' is being used to refer to the voluntary opening of a clenched fist. Bowerman holds that there is no obvious competitor to '*unsqueeze.' However, when presented with this concrete example, most native speakers will say that both 'release' and 'let go' are reasonable alternatives. The Competition Model claim is that, because there is no rote auditory support for '*unsqueeze,' forms like 'release' or 'let go' will eventually compete against and eliminate this particular error.

Several semantic cues support this process of recovery. In particular, inanimate objects such as rubber balls and sponges cannot be '*unsqueezed' in the same way that they can be 'squeezed.' Squeezing is only reversible if we focus on the action of the body part doing the squeezing, not the object being squeezed. It is possible that, at first, children do not fully appreciate these constraints on the reversibility of this particular action. However, it is equally likely that they resort to using '*unsqueeze' largely because of the unavailability of more suitable competitors such as 'release.' An error of this type is equivalent to production of 'falled' when the child is having trouble remembering the correct form 'fell.' Or consider the

competition between '*unapproved' and its acceptable competitor 'disapproved'. We might imagine that a mortgage loan application that was initially approved could then be subsequently '*unapproved.' We might have some uncertainty about the reversibility of the approval process, but the real problem is that we have not sufficiently solidified our notion of 'disapproved' in order to have it apply in this case. The flip side of this coin is that many of the child's extensional productions of reversives will end up being acceptable. For example, the child may produce 'unstick' without ever having encountered the form in the input. In this case, the form will survive. Although it will compete with 'remove', it will also receive occasional support from the input and will survive long enough for it to begin to carve out further details in the semantic scope of verbs that can be reversed with the prefix 'un-' (Li & MacWhinney, 1996).

### 4.4.2 Lexical competition

The same logic that can be used to account for recovery from morphological overgeneralizations can be used to account for recovery from lexical overgeneralizations. For example, a child may overgeneralize the word 'kitty' to refer to tigers and lions. The child will eventually learn the correct names for these animals and restrict the overgeneralized form. The same three forces are at work here: analogic pressure, competition, and episodic encoding. Although the child has never actually seen a 'kitty' that looks like a tiger, there are enough shared features to license the generalization. If the parent supplies the name 'tiger.' there is a new episodic encoding that then begins to compete with the analogic pressure. If no new name is supplied, the child may still begin to accumulate some negative evidence, noting that this particular use of 'kitty' is not yet confirmed in the input.

Merriman (1999) has shown how the linking of competition to a theory of attentional focusing can account for the major empirical findings in the literature on Mutual Exclusivity (the tendency to treat each object as having only one name). By treating this constraint as an emergent bias, we avoid a variety of empirical problems. Since competition is probabilistic, it only imposes a bias on learning, rather than a fixed innate constraint. The probabilistic basis for competition allows the child

to deal with hierarchical category structure without having to enforce major conceptual reorganization. Competition may initially lead a child to avoid referring to a 'robin' as a 'bird,' since the form 'robin' would be a better direct match. However, sometimes 'bird' does not compete directly with 'robin.' This occurs when referring to a collection of different types of birds that may include robins, when referring to an object that cannot be clearly identified as a robin, or when making anaphoric reference to an item that was earlier mentioned as a 'robin.'

### 4.4.3 Syntactic frame competition

Overgeneralizations in syntax arise when a feature-based construction common to a group or 'gang' of verbs is incorrectly overextended to a new verb. This type of overextension has been analyzed in both distributed networks (Miikkulainen & Mayberry, 1999) and interactive activation networks (Elman *et al.*, 2005; MacDonald *et al.*, 1994; MacWhinney, 1987b). These networks demonstrate the same gang effects and generalizations found in networks for morphological forms (Plunkett & Marchman, 1993) and spelling correspondences (Taraban & McClelland, 1987). If a word shares a variety of semantic features with a group of other words, it will be treated syntactically as a member of the group.

Consider the example of overgeneralizations of dative movement. Verbs like 'give', 'send', and 'ship' all share a set of semantic features involving the transfer of an object through some physical medium. In this regard, they are quite close to a verb like 'deliver' and the three-argument verb group exerts strong analogic pressure on the verb 'deliver'. However, dative movement only applies to certain frequent, monosyllabic transfer verbs and not to multisyllabic, Latinate forms with a less transitive semantics such as 'deliver' or 'recommend.' When children overgeneralize and say, 'Tom delivered the library the book,' they are obeying analogic pressure from the group of transfer verbs that permit dative movement. In effect, the child has created a new argument frame for the verb 'deliver.' The first argument frame only specifies two arguments – a subject or 'giver' and an object or 'thing transferred.' The new lexical entry specifies three arguments. These two homophonous entries for 'deliver' are now in competi-

tion, just as '*goed' and 'went' were in competition. Like the entry for '*goed', the three-place entry for 'deliver' has good analogic support, but no support from episodic encoding derived from the input. Over time, it loses in its competition with the two-argument form of 'deliver' and its progressive weakening along with strengthening of the competing form leads to recovery from overgeneralization. Thus, the analysis of recovery from 'Tom delivered the library the book' is identical to the analysis of recovery from '*goed'.

### 4.4.4 Modeling construction strength

It may be useful to characterize the temporal course of competitive item-based learning in slightly more formal terms. To do this, we can say that a human language is generated by the application of a set of constructions that map arguments to predicates. For each item-based construction (IC), there is a correct mapping (CM) from argument to its predicates and any number of incorrect mappings (IM). The IMs receive support from analogical relations to groups of CM with similar structure. From these emerge feature-based constructions (FC). The CMs receive support from positive input, as well as analogical relations to other CMs and FCs. Each positive input increases the strength S of a matching CM by amount A. Learning of an IC occurs when the S of CM exceeds the S of each of the strongest competing IM by some additional amount. This is the dominance strength or DS.

To model language learning within this framework, we need to understand the distribution of the positive data and the sources of analogical support. From database searches and calculation of ages of learning of CM, we can estimate the number of positive input examples (P) needed to bring a CM to strength DS. For each C, if the input has included P cases by time T, we can say that a particular CM reaches DS monotonically in time T. At this point, IC is learned. Languages are learnable if their component ICs can be learned in time T. To measure learning to various levels, we can specify learning states in which there remain certain specified slow constructions (SC) that have not yet reached DS. Constructions learned by this time can be called NC or normal constructions. Thus, at time T, the degree of completion of the learning of L can be expressed as NC/NC + SC.

This is a number that approaches 1.0 as T increases. The residual presence of a few SC, as well as occasional spontaneous declines in DS of CM will lead to deviations from 1.0. The study of the SCs requires a model of analogic support from FCs. In essence, the logical problem of language acquisition is then restated as the process of understanding how analogical pressures lead to learning courses that deviate from what is predicted by simple learning on positive exemplars for individual item-based constructions.

### 4.5 Cue construction

The fifth solution to the logical problem and the second of the solutions that promotes recovery from overgeneralization is cue construction. Most recovery from overgeneralization relies on competition. However, competition will eventually encounter limits in its ability to deal with the fine details of grammatical patterns. To illustrate these limits, consider the case of recovery from resultative overgeneralizations such as '*I untied my shoes loose'. This particular extension receives analogic support from verbs like 'shake' or 'kick' which permit 'I shook my shoes loose' or 'I kicked my shoes loose.' It appears that the child is not initially tuned in to the fine details of these semantic classifications. Bowerman (1988) has suggested that the process of recovery from overgeneralization may lead the child to construct new features to block overgeneralization. We can refer to this process as 'cue construction.'

Recovering from other resultative overgeneralizations may also require cue construction. For example, an error such as '*The gardener watered the tulips flat' can be attributed to the operation of a feature-based construction which yields three-argument verbs from 'hammer' or 'rake', as in 'The gardener raked the grass flat.' Source-goal overgeneralization can also fit into this framework. Consider, '*The maid poured the tub with water' instead of 'The maid poured water into the tub' and '*The maid filled water into the tub' instead of 'The maid filled the tub with water.' In each case, the analogic pressure from one group of words leads to the establishment of a case frame that is incorrect for a particular verb. Although this competition could be handled just by the strengthening of the correct patterns, it seems likely that the child

also needs to clarify the shape of the semantic features that unify the 'pour' verbs and the 'fill' verbs.

Bowerman (personal communication) provides an even more challenging example. One can say 'The customers drove the taxi driver crazy,' but not '*The customers drove the taxi driver sad.' The error involves an overgeneralization of the exact shape of the resultative adjective. A connectionist model of the three-argument case frame for 'drive' would determine not only that certain verbs license a third possible argument, but also what the exact semantic shape of that argument can be. In the case of the standard pattern for verbs like 'drive,' the resultant state must be terminative, rather than transient. To express this within the Competition Model context, we would need to have a competition between a confirmed three-argument form for 'drive' and a looser overgeneral form based only on analogic pressure. A similar competition account can be used to account for recovery from an error such as, '*The workers unloaded the truck empty' which contrasts with 'The workers loaded the truck full'. In both of these cases, analogic pressure seems weak, since examples of such errors are extremely rare in the language learning literature.

The actual modelling of these competitions in a neural network will require detailed lexical work and extensive corpus analysis. A sketch of the types of models that will be required is given in MacWhinney (1999).

## 4.6 Monitoring

The sixth solution to the logical problem involves children's abilities to monitor and detect their own errors. The Competition Model holds that, over time, correct forms gain strength from encounters with positive exemplars and that this increasing strength leads them to drive out incorrect forms. If we make further assumptions about uniqueness, this strengthening of correct forms can guarantee the learnability of language. However, by itself, competition does not fully account for the dynamics of language processing in real social interactions. Consider a standard self-correction such as 'I gived, uh, gave my friend a peach.' Here the correct form 'gave' is activated in real time just after the production of the overgeneralization. MacWhinney (1978) and Elbers & Wijnen (1993) have treated this type of self-correction as involving 'expressive monitoring' in which the child listens to her own output, compares the correct weak rote form with the incorrect overgeneralization, and attempts to block the output of the incorrect form. One possible outcome of expressive monitoring is the strengthening of the weak rote form and weakening of the analogic forms. Exactly how this is implemented will vary from model to model

In general, retraced false starts move from incorrect forms to correct forms, indicating that the incorrect forms are produced quickly, whereas the correct rote forms take time to activate. Kawamoto (1994) has shown how a recurrent connectionist network can simulate exactly these timing asymmetries between analogic and rote retrieval. For example, Kawamoto's model captures the experimental finding that incorrect regularized pronunciations of 'pint' to rhyme with 'hint' are produced faster than correct irregular pronunciations.

An even more powerful learning mechanism is what MacWhinney (1978) called 'receptive monitoring.' If the child shadows input structures closely, he will be able to pick up many discrepancies between his own productive system and the forms he hears. Berwick (1987) found that syntactic learning could arise from the attempt to extract meaning during comprehension. Whenever the child cannot parse an input sentence, the failure to parse can be used as a means of expanding the grammar. The kind of analysis through synthesis that occurs in some parsing systems can make powerful use of positive instances to establish new syntactic frames. Receptive monitoring can also be used to recover from overgeneralization. The child may monitor the form 'went' in the input and attempt to use his own grammar to match that input. If the result of the receptive monitoring is '*goed', the child can use the mismatch to reset the weights in the analogic system to avoid future overgeneralizations.

Neural network models that rely on back-propagation assume that negative evidence is continually available for every learning trial. For this type of model to make sense, the child would have to depend heavily on both expressive and receptive monitoring. It is unlikely that these two mechanisms operate as continuously as would be required for a mechanism such as back-propagation. However, not all connectionist models rely on the availability of negative evidence. For example, Kohonen's self-organizing feature map model

(Miikkulainen, 1993) learns linguistic patterns simply using cooccurences in the data with no reliance on negative evidence.

## 4.7  Indirect negative evidence

The seventh solution to the logical problem of language acquisition relies on the computation of indirect negative evidence. This computation can be illustrated with the error '*goed.' To construct indirect negative evidence in this case, children need to track the frequency of all verbs and the frequency of the past tense as marked by the regular '-ed.' Then they need to compute regular '-ed' as a percentage of all verbs.  Next they need to track the frequency of the verb 'go' in all of its uses and the frequency of '*goed''. To gain a bit more certainty, they should also calculate the frequency of a verb like 'jump' and the frequency of 'jumped.'  With these ratios in hand, the child can then compare the ratio for 'go' with those for 'jump' or verbs in general and conclude that the attested cases of '*goed' are fewer than would be expected on the basis of evidence from verbs like 'jump.'  They can then conclude that '*goed' is ungrammatical. Interestingly, they can do this without receiving overt correction.

The structures for which indirect negative evidence could provide the most useful accounts are ones that are learned rather late. These typically involve low-error constructions of the type that motivate the strong form of the logical problem. For example, children could compute indirect negative evidence that would block wh-raising from object-modifying relatives in sentences such as (37).

36. The police arrested the thieves who were carrying the loot.
37. *What did the police arrest the thieves who were carrying?
38. To do this, they would need to track the frequency of sentences such as:
39. Bill thought the thieves were carrying the loot.
40. What did Bill think the thieves were carrying?

Noting that raising from predicate complements occurs fairly frequently, children could reasonably conclude that the absence of raising from object modification position means that it is ungrammatical. Coupled with conservatism, indirect negative evidence can be a useful mechanism for avoiding overgeneralization of complex syntactic structures.

The item-based acquisition component of the Competition Model provides a framework for computing indirect negative evidence. The indirect negative evidence tracker could note that, although 'squeeze' occurs frequently in the input, '*unsqueeze' does not. This mechanism works through the juxtaposition of a form receiving episodic support ('squeeze') with a predicted inflected form ('unsqueeze').

This mechanism uses analogic pressure to predict the form '*unsqueeze.' This is the same mechanism as used in the generation of '*goed.' However, the child does not need to actually produce '*unsqueeze,' only to hypothesize its existence. This form is then tracked in the input. If it is not found, the comparison of the near-zero strength of the unconfirmed form 'unsqueeze' with the confirmed form 'squeeze' leads to the strengthening of competitors such as 'release' and blocking of any attempts to use 'unsqueeze.' Although this mechanism is plausible, it is more complicated than the basic competition mechanism and places a greater requirement on memory for tracking of non-occurrences. Since the end result of this tracking of indirect negative evidence is the same as that of the basic competition mechanism, it is reasonable to imagine that learners use this mechanism only as a fall back strategy, relying on simple competition to solve most problems requiring recovery from overgeneralization.

## 5.  Consequences and Conclusions

This analysis suggests that we should not longer speak of language learning as being confined by the poverty of positive evidence or negative evidence. Both types of evidence are far more abundant than has been imagined. Nor should we assume that recovery from overgeneralization involves a fundamental logical problem. Recovery is supported by a set of four powerful processes (competition, cue construction, monitoring, and indirect negative evidence) that provide redundant and complementary solutions to the logical problem. In addition, we know that alternative characterizations of the nature of the target grammar can

take much of the logical bit out of the logical problem. Finally, we have seen that the language addressed to children is not at all unparsable or degenerate, once a few superficial retracing structures are repaired.

We have reviewed seven solutions to the logical problem that work together to buffer the process of language acquisition. When we consider the interaction of the seven solutions in this way, we soon come to realize the pivotal role played by the item-based construction. First, the item-based construction directly enforces conservatism by requiring that each generalization of each argument frame be based on directly observable positive evidence. Second, the probabilistic competition between item-based constructions provides a meaningful way of understanding the probabilistic nature of grammar. Third, the competition between item-based constructions directly promotes recovery from overgeneralization. Fourth, the additional mechanisms of cue construction, indirect negative evidence, and monitoring serve to fine-tune the operations of competition. These processes operate particularly in those cases where uniqueness is not fully transparent or where the restriction of a general process requires additional fine-tuning of cues.

The current analysis assigns great importance to good positive data. Marcus (1993) has suggested that parents are inconsistent in their provision of negative evidence to the child. But the Competition Model assumes that it is positive data that is crucial for learning. One way in which a parent can provide crucial positive evidence is through recasting, but other methods are possible too. In various cultures and subgroups, positive evidence can be presented and focused through elicited repetition, choral recitation of stories, interaction with siblings, or games. Methods that emphasize shared attention and shared understanding can guide children toward the control of literate expression. This shared attention can arise in groups of co-wives in Central Africa just as easily as it can from isolated mother–child dyads in New England.

Recently, Hauser, Chomsky, & Fitch (2002) have argued that the core evolutionary adaptation that was required to support human language involved the introduction of a facility for recursion. The analysis in the current paper modifies and extends this claim by emphasizing the evolutionary (MacWhinney, 2005) and developmental (Tomasello, 2000) centrality of the item-based construction as the controller of recursive composition of phrases and sentences. However MacWhinney (2005) views linguistic recursion as emerging gradually from preexisting structures in spatial cognition, rather than as appearing suddenly during the Late Pleistocene. Studies of the functional neural underpinnings of recursion can go a long ways toward clarifying the details of these issues.

## References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *The emergence of language* (pp. 359-380). Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533-581.

Berwick, R. (1987). Parsability and learnability. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bohannon, N., MacWhinney, B., & Snow, C. (1990). No negative evidence revisited: Beyond learnability or who has to prove what to whom. *Developmental Psychology, 26*, 221-226.

Bowerman, M. (1987). Commentary. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bowerman, M. (1988). The "no negative evidence" problem. In J. Hawkins (Ed.), *Explaining language universals* (pp. 73-104). London: Blackwell.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 11-54). New York: Wiley.

Buttery, P. (2004). A quantitative evaluation of naturalistic models of language acquisi**tion; the** efficiency of the Triggering Learning Algorithm compared to a Categorial Grammar Learner. *Coling 2004*, 1-8.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1981). *Lectures on government and binding*. Cinnaminson, NJ: Foris.

Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In J. Jacobs (Ed.), *Syntax: An international handbook of contemporary research* (pp. 1-32). Berlin: Walter de Gruyter.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63 No. 3*, 522-543.

De Villiers, J., Roeper, T., & Vainikka, A. (1990). The acquisition of long distance rules. In L. Frazier & J. De Villiers (Eds.), *Language processing and language acquisition*. Amsterdam: Kluwer.

Elbers, L., & Wijnen, F. (1993). Effort, production skill, and language learning. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 337-368). Timonium, MD: York.

Elman, J. L., Hare, M., & McRae, K. (2005). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fodor, J., & Crain, S. (1987). Simplicity and generality of rules in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. Hillsdale, N.J.: Lawrence Erlbaum.

Gold, E. (1967). Language identification in the limit. *Information and Control, 10*, 447-474.

Hauser, M., Chomsky, N., & Fitch, T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Hausser, R. (1999). *Foundations of computational linguistics: Man-machine communication in natural language*. Berlin: Springer.

Hopcroft, J., & Ullman, J. (1979). *Introduction to automata theory, languages, and computation*. Reading, Mass.: Addison-Wesley.

Horning, J. J. (1969). *A study of grammatical inference*: Stanford University, Computer Science Department.

Hornstein, N., & Lightfoot, D. (1981). *Explanation in linguistics: the logical problem of language acquisition*. London: Longmans.

Jain, S., Osherson, D., Royer, J., & Sharma, A. (1999). *Systems that learn*. Cambridge, MA: MIT Press.

Jespersen, O. (1922). *Language: Its nature, development, and origin*. London: George Allen and Unwin.

Kanazawa, M. (1998). *Learnable classes of categorial grammars*. Stanford, CA: CSLI Publications.

Kawamoto, A. (1994). One system or two to handle regulars and exceptions: How time-course of processing can inform this debate. In S. D. Lima, R. L. Corrigan & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 389-416). Amsterdam: John Benjamins.

Kimball, J. (1973). *The formal theory of grammar*. Englewood Cliffs, NJ: Prentice-Hall.

Lewis, J. D., & Elman, J. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the 26th Annual Boston University Conference on Language Development*.

Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science, 8*, 3-30.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676-703.

MacWhinney, B. (1975). Pragmatic patterns in child syntax. *Stanford Papers And Reports on Child Language Development, 10*, 153-165.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43*, Whole no. 1, pp. 1-123.

MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Vol. 1. Syntax and semantics* (pp. 73-136). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1987a). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1987b). Toward a psycholinguistically plausible parser. In S. Thomason (Ed.), *Proceedings of the Eastern States Conference on Linguistics*. Columbus, Ohio: Ohio State University.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2005). Language evolution and human development. In D. Bjorklund & A. Pellegrini (Eds.), *Child development and evolutionary psychology*. New York: Academic.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 29*, 121-157.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*, 255-277.

Maratsos, M., Kuczaj, S. A., Fox, D. E., & Chalkley, M. A. (1979). Some empirical studies in the acquisition of transformational relations: Passives, negatives, and the past tense. In W. A. Collins (Ed.), *Children's language and communication*. Hillsdale, N.J.: Lawrence Erlbaum.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition, 46*, 53-85.

Merriman, W. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp. 331-358). Mahwah, NJ: Lawrence Erlbaum.

Miikkulainen, R. (1993). *Subsymbolic natural language processing*. Cambridge, MA: MIT Press.

Miikkulainen, R., & Mayberry, M. R. (1999). Disambiguation and grammar as emergent soft constraints. In B. MacWhinney (Ed.), *The emergence of language* (pp. 153-176). Mahwah, NJ: Lawrence Erlbaum Associates.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. Snow (Ed.), *Talking to children: Language input and acquisition*. Cambridge: Cambridge University Press.

O'Grady, W. (1997). *Syntactic development*. Chicago: Chicago University Press.

Piattelli-Palmarini, M. (1980). *Language and learning: the debate between Jean Piaget and Noam Chomsky*. Cambridge MA: Harvard University Press.

Pinker, S. (1994). *The language instinct*. New York: William Morrow.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building. *Cognition, 49*, 21-69.

Pullum, G., & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review, 19*, 9-50.

Reich, P. (1969). The finiteness of natural language. *Language, 45*, 831-843.

Sagae, K., MacWhinney, B., & Lavie, A. (2004). Adding syntactic annotations to transcripts of parent-child dialogs. In *LREC 2004* (pp. 1815-1818). Lisbon: LREC.

Sakas, W., & Fodor, J. (2001). The structural triggers learner. In S. Bertolo (Ed.), *Language acquisition and learnability*. New York: Cambridge University Press.

Saxton, M. (1997). The Contrast Theory of negative input. *Journal of Child Language, 24*, 139-161.

Shinohara, T. (1994). Rich classes inferable from positive data: length-bounded elementary formal systems. *Information and Computation, 108*, 175-186.

Taraban, R., & McClelland, J. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*, 608-631.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156-163.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*, 23-79.

Wexler, K., & Hamburger, H. (1973). On the insufficiency of surface data for the learning of transformational languages. In K. Hintikka (Ed.), *Approaches to natural language*. Dordrecht-Holland: D. Reidel.

Wilson, B., & Peters, A. M. (1988). What are you cookin' on a hot? Movement Constraints in the Speech of a Three-Year-Old Blind Child. *Language, 64, No.2*, 249-273.

Wolfe Quintero, K. (1992). Learnability and the acquisition of extraction in relative clauses and wh-questions. *Studies in Second Language Acquisition, 14*, 39-70.