A Survey on Computational Models for Argument Structure Acquisition

Afra Alishahi
Department of Computer Science
University of Toronto

June 7, 2005

Contents

1 Overview

In most natural languages, verbs are used to describe actions and situations. Often verbs have several possible patterns of arguments, or *argument structures*, which indicate the syntactic position of arguments and the thematic roles assigned to them. The variation involves not only the number and syntactic type of arguments that a verb can take, but also the allowable combinations of these arguments.

The acquisition of verb argument structure in children is a complex phenomenon, generating much debate. As early as age three, children seem to possess some knowledge about the general thematic linking rules, i.e., the regularities in the relationship between thematic roles such as agent and patient and syntactic functions such as subject and direct object (MacWhinney, 1995; Demuth et al., 2002). They use this knowledge to produce novel utterances they have never heard before by generalizing the argument structure of verbs they have already learned to new ones. This ability sometimes leads to overgeneralization, e.g., **I said her no* or **you can drink me the milk* (Bowerman, 1982, 1996). These kinds of errors gradually cease as children grow older, and by their teenage years they have acquired almost adult-like linguistic competence (Demuth et al., 2002; Bowerman, 1996).

The way children use language regularities to generalize to new forms reveals a great deal about the internal mechanisms of argument structure acquisition in human beings. Evidence from spontaneous speech data shows consistent cross-linguistic patterns of generalization for a variety of linguistic contexts such as morphological inflections and syntactic constructions (Marcus et al., 1992; Demuth et al., 2003; Bowerman, 1982). The process of mastering the above-mentioned regularities in verb argument structure is noteworthy especially when we consider the limitations governing the acquisition process, the most important of which is the lack of negative evidence (Marcus, 1993). By negative evidence, we refer to corrective feedback from caretakers in response to children's speech. Positive evidence, on the other hand, refers to the "correct" linguistic input the child amply receives from the environment. Any theory of argument structure acquisition (and language acquisition in general) should explain how children recover from making errors (especially generalization errors) only by receiving positive evidence.

Among the numerous hypotheses proposed for explaining different aspects of verb argument structure acquisition, one can distinguish two major camps: nativist theories, which state that children are helped by means of their innate knowledge about the language, and usage-based theories, which argue that children learn all they know about verb argument structure from the input they receive. Research evaluating these theories by testing them against spontaneous speech data gathered from children learning language (e.g., (Bowerman, 1982, 1996; Demuth et al., 2002)) usually favours the usage-based approach.

In an attempt to simulate the argument structure acquisition process and evaluate the proposed hypotheses, a number of computational models have been developed. Connectionist models (Allen, 1997; Allen and Seidenberg, 1999; Jones et al., 2000; Desai, 2002) have been especially popular because they reflect the emergence of abstract knowledge from exposure to a number of instances. But they are usually simplistic and not easily extendable; moreover, few of these models can *use* the acquired knowledge, and so the study of generalization patterns is limited. Other models such as those of Clark (2001b), Gobet et al. (2004), Solan et al. (2004) and Cartwright and Brent (1997) aim for learning syntactic knowledge from corpus data. These models ignore the role of semantics in learning syntax and the relationship between the two, and most of them violate cognitive plausibility criteria such as incrementality. However, a number of computational models have been presented in recent years that receive pairs of form and meaning as input. These models pursue a variety of goals, such as learning lexical constructions (Chang, 2004), learning verb meaning from image data (Dominey, 2003; Dominey and Inui, 2004), acquiring the syntactic and semantic features of verbs (Niyogi, 2002) and setting parameters of a universal grammar (Buttery, 2003, 2004). However, the motivation behind all of them is to explore the relationship between the external form (syntax) and the meaning (semantics) for the purpose of language acquisition.

In the next section, some results of analyzing the available psycholinguistic data are examined, including overgeneralization patterns and the lack of negative evidence. The following section reviews and compares some important theories of verb argument structure acquisition based on the available data. Some of the existing computational models that simulate the process of acquisition based on different theories are then reviewed.

2 Psycholinguistic findings

Many issues arise in the experimental data about children learning language; here we focus on two which are most relevant to the acquisition of verb argument structure: overgeneralization patterns, and the lack of negative evidence.

2.1 Overgeneralization patterns

Children freely generalize regular grammatical patterns to new terms they learn or even to old ones. This ability sometimes leads to overgeneralized forms, such as *we goed* and *Adam fall toy* (MacWhinney, 1995). In order to discover the detailed patterns of generalization and overgeneralization, Marcus et al. (1992) provide quantitative data on overregularization of English past tense verbs. The analysis of the data reveals some interesting facts. First, an extended period of correct performance precedes the first overregularization errors. Before that point (around age two), a reliable correlation between verb tense and the regular morphological marking has not yet been detected and established. Second, overregularizations occur at a low, almost constant rate into the school age years. Third, overregularization is influenced by the frequency of the irregular form, i.e., irregular forms frequently used by parents are less often overregularized. Fourth, parameters such as the child's vocabulary size or the number or proportion of irregular verbs to regular verbs do not have a significant correlation with the overregularization rate.

Marcus et al. (1992) propose an explanation for these findings: children, like adults, mark tense using both memorization of particular forms and an affixation rule that can generate a regular past tense form for any verb. Retrieval of an irregular blocks the rule, but children's memory links are not strong enough to guarantee perfect retrieval. When retrieval fails, the rule is applied, and overregularization happens. This explanation may work for cases such as the English past tense where there are a few dominant rules and a number of exceptions for each. However, the case of overregularization in syntactic constructions, in which there are a number of competing syntactic forms for each verb, is not easily explained this way.

The study of overgeneralization in the argument structure of causative verbs (Bowerman, 1982, 1996), as well as of verbs with theme and location/goal/source arguments and three-argument verbs with recipient or beneficiary roles (Bowerman, 1990), reveals a similar pattern to what Marcus et al. (1992) show for irregular past tense. Bowerman (1982) uses spontaneous speech records from her two daughters from ages one to six to study overregularizations involving the creation of novel causative verbs. She shows that children begin to create novel causative verbs between the ages of about two and three, using a predicate that is normally noncausative to convey a causative meaning, such as *I'm just gonna fall this on her*. The onset of the errors is preceded by a period of several months in which the relevant verbs have already been consistently used in a syntactically appropriate way. The errors continue to occur through teenage years,[1] and finally become relatively infrequent by adulthood. She also reports a striking finding: the first novel causative verbs do not appear until several months after the first transitive uses of legitimate causative verbs like *open*, and they appear almost simultaneously with the onset of the first periphrastic causatives with *make* and *get*. She suggests that the temporal correspondences between the emergence of periphrastic causatives and novel lexical causatives can be seen as the result of the child's having grasped the structure that underlies both.

2.2 Lack of negative evidence

Marcus (1993) examines data provided by other researchers in support of or against the hypothesis that children rely on negative evidence from their parents to detect errors in their speech and correct it. He shows that the only proven type of negative evidence available to children is noisy feedback, which he defines as corrective feedback provided for some ungrammatical sentences as well as for some grammatical ones. Therefore, a single instance of reply cannot guarantee that a given sentence is ungrammatical. To use noisy feedback effectively, the child must first figure out the probabilities of reply instances for both grammatical and ungrammatical sentences, which necessitates repeating each sentence an unrealistically large number of times.

Marcus defines four criteria to determine whether a particular type of noisy feedback could be necessary for unlearning errors: (1) the reply type must be available to all children; (2) the reply type must be available throughout acquisition; (3) the reply type

[1] As an example of overgeneralization in teenage years, one child (13 years, 8 months), talking about a movie, said *It's not really scary, but some things jump you* (meaning *some things make you jump*). It was the first time he had used *jump* as a causative verb (Suzanne Stevenson, personal communication).

must be available for errors in each component of language; (4) the reply type must be available for all types of errors within a given component. Any pattern of noisy feedback that is necessary for language acquisition should meet all these criteria. But there is no evidence that any type of parental feedback is available widely enough to fulfill this requirement. Positive feedback, on the other hand, meets all of the above criteria. Thus, the problem of accounting for children's avoidance of and recovery from errors in language acquisition is likely to be explained by the nature of their internal learning mechanisms which rely on positive evidence.

3 Theories of verb argument structure acquisition

Various proposals have been put forward to account for the acquisition of verb argument structure. Nativist approaches, such as Chomsky (1986), Pinker (1984, 1989) and Gleitman (1990), assume that the acquisition process is facilitated by means of children's innate knowledge about language. Usage-based approaches such as Tomasello (2000a), Tomasello and Abbot-Smith (2002), Goldberg (1999) and Bowerman (1990, 1996), on the other hand, claim that children learn all they know about verb argument structure from the input they receive.

We will review some pioneering hypotheses from each approach in the following subsections. From the nativist camp, we briefly review Chomsky's theory of Universal Grammar and its main argument, the Poverty of Stimulus, and Pinker's Semantic Bootstrapping Hypothesis. From the usage-based camp, we review Tomasello's Verb Island Hypothesis, and Goldberg's Argument Structure Constructions. We then evaluate each theory against available psycholinguistic data, and compare the success of the predictions each theory makes based on the data.

3.1 Nativist theories

In the nativist view of language acquisition, children are assumed to be equipped with innate knowledge of language. The extent of this knowledge, however, is not agreed upon among all the researchers in this group. The innateness theory, first proposed by Chomsky (1986), states that children do not have to learn or construct abstract syntactic structures, but rather they already possess them as part of their innate language faculty.

Pinker (1984, 1989) specialized this theory to verb argument structure acquisition under the name of the Semantic Bootstrapping Hypothesis. We will look at each of these two hypotheses in more detail.

3.1.1 Chomsky: Universal Grammar

Chomsky (1986) believes that there is a particular component of the human mind, the *language faculty*, which determines the form and meaning of expressions of the language. This language faculty has encoded in it the general principles and elements common to all human languages, or *Universal Grammar* (UG). The principal argument for the innateness of UG is *the Argument from the Poverty of Stimulus* (APS), which states that there are principles of grammar that cannot be learned on the basis of positive input alone, however complete and grammatical that evidence is.

The APS rests on the premise that the knowledge of syntax in the human brain has the form of a universal grammar. To be descriptive enough to account for all the complexities of all existing natural languages, universal grammar must be very rich in structure. Also, to account for the variety of languages, the grammar must be augmented by a large number of parameters to allow for flexibility. This way, the acquisition of a language from data is equivalent to finding the right set of parameters that best explains the data. Knowing the number of parameters required and the number of possible values for each, the problem can be formalized as a search problem in a huge search space. In this framework, the data available to children is not sufficient on its own for learning language. However, the validity of the APS depends on the assumed representation for the knowledge of language, that is, universal grammar. Besides, the role of distributional statistics and probabilistic reasoning is completely neglected in this theory.

3.1.2 Pinker: Semantic Bootstrapping Hypothesis

The semantic bootstrapping hypothesis, proposed by Pinker (1984), is a nativist account of verb argument structure acquisition. According to Pinker, early acquisition of verb argument structure is regulated by a *canonical mapping scheme* — that is, a default mapping between thematic roles and syntactic functions such that the subject of the sentence is mostly associated to the agent role, the direct object, to the theme/patient role, and the oblique object, to the location/goal/source role. In this account, children are assumed to apply the canonical mapping scheme to new predicate-argument sequences. That is, they link the first function on the syntactic hierarchy to the first available role on the thematic hierarchy, and then move along to the next syntactic function, until all syntactic functions are linked to appropriate thematic roles. Some predicates or constructions assign thematic roles to syntactic functions non-canonically, i.e., they do not obey the above order. Children must learn such cases by observing how adults treat the predicate syntactically. The irregular mapping of such predicates blocks the default application of canonical linking.

The proposed mapping scheme is assumed to be universal, and to reflect properties of children's innate capacity for language acquisition. An important problem with this theory is that linguists do not agree on what constitutes the canonical mapping between thematic roles and syntactic functions. It seems that there is no cross-linguistically consistent relationship between thematic and syntactic roles; for example, in some languages the patient is linked to subject and the agent is linked to the object (Dixon, 1972; Marantz, 1984; cited by Bowerman, 1990). A second problem is that, before children can refer to their linking hierarchies to decide how to handle the arguments of a predicate syntactically, they have to know how many arguments the predicate has and what their thematic roles are. There are important cross-linguistic differences in the argument structures of the predicates that a child would hear in a given context, and even within a language there is ambiguity regarding the argument structure of verbs, which makes the problem more complicated.

Pinker (1989) later revised his canonical-mapping model in favour of an alternative model, according to which all verbs are consistent with linking rules. In the new theory, instead of thematic roles being ordered in a fixed hierarchy, they are positions in decompositional representations of verbs' meanings; for example, agent is the first argument of CAUSE, patient is the second argument of CAUSE, and theme is the first argument of GO and BE. This way, each thematic role is associated with its own linking rule (e.g., theme is linked to subject if that syntactic function has not already been assigned, otherwise to object). Since this model specifies that the way a verb's arguments are linked follows directly from its semantic representation, differences in the way closely related verbs (or alternative constructions involving the same verb) map their arguments must reflect differences in their meaning. Thus, since a verb's semantic structure and the mapping of its arguments are in perfect correspondence, a child with innate linking rules can predict correct mappings once he knows what a verb means.

Since the linking rules proposed by the new model are simpler and more general and apply equally to all verbs, they provide much less precise guidelines for the semantic bootstrapping of phrase-structure rules and for predicting the subcategorization frames of newly acquired verbs. This means that the burden is on the child to learn these highly articulated semantic structures. Besides, even if these general rules prove to be useful in acquiring verb argument structure and avoiding overregularization errors, it is not necessary to assume that the linking rules Pinker invokes are innate.

3.2 Usage-based theories

As mentioned before, the nativist view of language acquisition is mostly based on the arguments against the possibility of learning all the complexities of a natural language from the available input. This view assumes that human linguistic knowledge is characterized in terms of a Chomskyan universal grammar, and therefore learning a language is reduced to adjusting UG to the input data. However, a number of psycholinguistic experiments question this view.

Experimental studies on language comprehension and generation in young children (up to age three) show that they build their linguistic knowledge around individual items (Akhtar, 1999; Tomasello, 2000a; Bowerman, 1982), rather than adjusting some general grammar rules they already possess. Two-year old children show little tendency to apply syntactic structures they have already learned to new verbs, but rather conservatively use each verb in structures they have heard it in before. Also, young language learners quite readily substitute nominals for one another and begin to form an abstract category, "nominal", from very early in development, but are not productive with their language in some other ways. The tendency to generalize familiar constructions to new forms increases as the subjects of the studies grow older, showing that some general structures are learned over time.

These findings have led to the development of a number of usage-based theories of argument structure acquisition, which try to explain the emergence of abstract linguistic knowledge from input data. Two important theories of this group are reviewed here: Tomasello's Verb Island Hypothesis, and Goldberg's Argument Structure Constructions.

3.2.1 Tomasello: Verb Island Hypothesis

Tomasello (2000a,b, 2003) proposes a usage-based account of argument structure acquisition called the *Verb Island Hypothesis*. Based on experimental evidence that beginning learners do not use a verb in a sentence frame in which they have not heard it used before, he suggests that children do not possess verb-general argument structure constructions into which different verbs may be placed as needed, but rather they are working more concretely with verbs as individual lexical items whose syntactic behaviours must be learned one by one. The hypothesis is that children have an early period in which each of their verbs forms its own "island" consisting of verb-specific constructions with open nominal slots.

According to the hypothesis, children use cognitive and socio-cognitive processes such as imitation (reproducing the language adults produce for the same communicative function), analogy, and structure mapping (detecting both structural and functional similarities in utterances independent of the specific words involved) to gradually categorize the relational-syntactic structure of their various item-based constructions, and therefore become productive with their language in more adult-like ways. Three stages are mentioned: the formation of verb island constructions around each individual verb by hearing a number of utterances containing it; the emergence of more general constructions such as the simple transitive construction from similar verb island constructions; and the abstraction of higher order constructions such as the subject-predicate construction from a group of similar general constructions.

Although the hypothesis explains the conservative use of language by young children and the gradual emergence of the general structures over time, it is vague in describing exactly how the more general categories and schemes form. For example, it is emphasized that at the beginning, there are no syntagmatic categories such as subject and object or even agent and patient with which children are working, but rather such verb-specific things as 'hitter' and 'hittee' or 'sitter' and 'thing sat upon.' However, how these verb-specific roles metamorphose to general thematic roles and eventually to the abstract syntactic categories remains unexplained.

3.2.2 Goldberg: Argument Structure Constructions

Goldberg (1999) suggests another way of capturing the correspondence between form and meaning by positing abstract constructions that pair form with meaning, independently of the verbs that appear in them. She offers an account of how the meaning associated with argument structure constructions is acquired: argument structure patterns are initially acquired on a verb-by-verb basis, and constructions associated with the common syntactic patterns are learned through a process of categorization and generalization over the input. The generalization of constructional meaning is based largely on the meanings of highly frequent light verbs.[2] Children are likely to record a correlation between a certain formal pattern and the meaning of the particular verbs used earliest and most frequently in that pattern. Because light verbs are more frequent than other verbs and are learned early, they tend to be the ones around which constructional meaning centres.

In a later study, Bencini and Goldberg (2000) compare the role of the main verb and the argument structure construction in sentence meaning by asking a number of participants to sort sentences according to their meaning. The results suggest that adults probably see both verbs and constructions as relevant to establishing meaning. More contentful verbs (such as *throw* and *slice*) have a stronger effect on the meaning of a sentence than less contentful ones (such as *get* or *take*). On the other hand, some constructions (e.g., *ditransitive*) were easier to identify and influenced the meaning more than others.

Goldberg (1999) argues that the argument structure constructions are also responsible for the unusual meaning associated to verbs when they appear in constructions that are not typical for them, e.g. *the truck rumbled down the street* (meaning *the truck moved down the street, rumbling*). What is not answered by this theory is why some verbs *cannot* participate in a particular construction at all, e.g. **the girl died down the street* (meaning the girl went down the street dying).

3.3 Comparing different theories based on experimental data

In order to evaluate Pinker's (1984) canonical mapping hypothesis, Bowerman (1990) compares the syntactic handling of verbs with a theme argument as subject to verbs with a location/goal/source argument as subject in the language development of her two English-speaking daughters. The data provide no support for the hypothesis that

[2]Light verbs are highly frequent verbs with very general meanings, such as *go, do, make, give, put.* Goldberg (1995) shows that the meanings of some of the light verbs correspond closely to the meanings associated with argument structure constructions. For example, the intransitive construction "Subj V Obl" paired with the meaning"X moves Y" corresponds to the meaning of the light verb *go*.

children receive selective help from innate linking rules for verbs of one type over verbs of the other type. Another analysis of these children's acquisition of verbs that have an agent subject and allow either the item transferred or shown (theme) or the recipient (goal) to be direct object, supports the above conclusion that canonically linked verbs are no easier for children than non-canonically linked verbs. There is no advantage even for prototypical agent-patient verbs: the subjects learn them at the same time as they handle a variety of other verb types.

Pinker (1989) suggests that some verb meanings correspond directly to children's mental representations of the relevant events (i.e., there is only one candidate agent, patient, etc.), while others are more ambiguous (e.g., for *pour* and *fill*, it is not clear whether the theme or the goal should be understood as the entity causally acted upon). Children must resolve meaning ambiguities by observing multiple uses of non-cognitively transparent verbs in context. In other cases, however, children may simply learn the mapping of the verb's arguments directly from adult speech. Regardless of which technique the child uses to determine the mapping of non-cognitively transparent verbs, it is clear that, under Pinker's theory, these verbs will require more effort than verbs whose meanings are cognitively transparent. To test the hypothesis that linking rules are innate, Bowerman (1990) compares strings with cognitively transparent verbs and more ambiguous verbs with respect to both time of onset and consistency of mapping. She shows that in the early grammatical development of her two subjects, there was no advantage for prototypical agent-patient verbs over verbs of possession like *have* and *got*, verbs of perception like *see* and *hear*, and so on. Two-argument strings with verbs from all of these categories emerged simultaneously, and the children in fact made more argument-ordering errors with prototypical agent-patient verbs than with verbs of other kinds. This evidence sits squarely counter to the hypothesis that knowledge of linking rules is innate.

Demuth et al. (2003) investigates children's learning of verb argument structure in Sesotho, where verb semantics does not play a central role in the acquisition process, since verbs from almost all semantic classes appear in certain constructions. The authors focus on the double object applicative construction, where theme and benefactive follow the verb and the animacy of the objects determines their order. The authors conducted a forced choice elicited production experiment to examine children's and adults' knowledge of animacy effects. They tested different age groups of subjects using a set of sentence pairs, and asked the subjects to select the correct sentence. The results show that 4-year-old children performed significantly above chance, showing that even young children have some knowledge of the language-specific word order constraints. However, the study also found that most 8-year-olds were not yet consistent in their performance, compared to adults' almost errorless performance. Thus, despite the lack of semantic verb class effects, learning the word order restrictions on Sesotho double object applicatives appears to be a gradual process. This poses a challenge to those nativist approaches to language acquisition, which assume that once a parameter is set, it is used correctly thereafter.

Akhtar (1999) provides experimental evidence in English that demonstrates both the item-based learning strategy and the emergence of general structures over time. Children aged 2, 3 and 4 were taught novel actions for two novel verbs modeled in SOV or VSO orders. Recording spontaneous productions of these verbs and elicited responses to queries of *what happened?*, Akhtar found that the group of 2- and 3–year-olds matched SOV or VSO patterns roughly half the time and corrected order to SVO order roughly half the time. The fact that young children produced the modeled orders of the novel verbs so often demonstrates that they were willing to learn the argument structures of these verbs on an individual basis. At the same time, the fact that these same children corrected the order for these same verbs half the time to make it standard English SVO order demonstrates that they were also aware of a generalization over the instances they had already learned. The 4-year-old children rarely matched the modeled order and were much more likely to correct to SVO order.

4 Computational models

The development of computational models for language acquisition has been facilitated by the availability of on-line resources such as the CHILDES database (MacWhinney, 1995) and the WordNet database (Miller, 1990), and also by the availability of sufficient computational power to use these resources in computer simulations of increasing scale. However, there are yet serious limitations in using them on a realistic scale, the most important of which is the lack of linguistic and non-linguistic input data that resemble real data children use for learning language. However, even limited computational models provide a test-bed for evaluating current theories of language acquisition, and can provide insight into plausible representations and learning mechanisms.

The next section is an examination of the problem of mapping words to meaning, the most significant model of which is Siskind (1994, 1996). Learning the mapping between the basic elements of language (words and meanings) is necessary in order to examine the relationship between the syntactic and semantic structures. In the following sections, we investigate a number of models for learning grammar either from form-meaning pairs or from text. All these methods are statistical and use a variety of techniques to perform the task. However, their input, basic assumptions and goals are different, and not all of them are cognitively plausible.

A separate subsection is dedicated to the examination of connectionist models, although most of them have many characteristics in common with those reviewed in the other subsections. Connectionist models have been popular as an alternative to the traditional symbolic accounts to language acquisition and processing, especially because their architecture resembles the physiology of human brain, and because they learn from experience and show the ability to generalize to novel cases. They provide an effective mechanism for learning general rules together with their exceptions (Christiansen and Chater, 1999). However, current connectionist models typically use toy fragments of grammar and small vocabularies. Aside from raising the question of the viability of scaling up, this makes it difficult to provide a detailed fit with the empirical data.

Following the popularity of the universal grammar hypothesis, a number of parameter-setting algorithms and models for adjusting universal grammar to input data were proposed, such as the Triggering Learning Algorithm (Gibson and Wexler, 1994), the Structural Triggers Learner (Fodor, 1998; Sakas and Fodor, 2001), and the Bayesian Incremental Parameter Setting algorithm (Briscoe, 1999, 2000). As the main purpose of this section is to focus on usage-based models that deal with the relationship between syntax and semantics, we do not review this group of algorithms here, except for Buttery (2004) which exploits a form-meaning pairing in the input data.

4.1 Mapping words to meaning

Most of the computational models for mapping words to meaning are based on the notion of *cross-situational learning* (Pinker, 1989; Fisher et al., 1994), a learning strategy for finding word meanings that are consistent across multiple situations. The main idea is to find a set of possible meanings in each situation and intersect those sets across all situations in which a word occurs to determine the meaning for that word. One of the main problems concerning this strategy is what Siskind (1996) calls *referential uncertainty*, that children hear utterances in contexts where more than one thing could have been said, and they must figure out which of those things is the meaning of the utterance just heard.

Siskind (1994, 1996) proposes an algorithm for the acquisition of word-to-meaning mappings which addresses both the problem of disambiguating referential uncertainty to determine the correct meaning of each utterance, and the problem of deciding how to break that utterance meaning into parts to assign to each word in the utterance as its correct meaning. Siskind formalizes the problem as follows: the learner is presented with a corpus of utterances, each paired with a set of conceptual expressions that represent hypothesized utterance meanings. Each utterance is viewed as a non-ordered collection of word symbols. The goal of the learner is to find a lexicon that maps each word symbol that appears in the corpus to a set of conceptual expressions that represent the meaning of different senses of that word. The proposed algorithm uses a set of principles to constrain hypotheses about the meanings of utterances that contain those words, e.g., that all components of the meaning of an utterance are derived from the meanings of words in that utterance. The model tests the effectiveness of these principles for performing lexical acquisition by way of computational simulation. The algorithm was later extended to handle noise and homonymy. However, the algorithm does not use any syntactic properties of the language heard, or any well-formedness properties of the underlying conceptual structure.

Thompson and Mooney (1999) present a system, WOLFIE (WOrd Learning From Interpreted Examples), that acquires a semantic lexicon of word-meaning pairs from a corpus of sentences paired with semantic representations in the form of a labeled tree. The goal is to find a semantic lexicon consistent with these data. A batching and greedy acquisition algorithm is proposed based on the cross-situational learning hypothesis, but the underlying principles are slightly different from Siskind's. The system handles homonymy, synonymy, multi-word lexemes, and null meaning for some phrases. A few assumptions are made to constrain hypotheses about the meanings of utterances and words, such as compositionality and the absence of noise. A greedy search is then performed, at each step of which the best phrase-meaning pair is chosen and added to the lexicon. A heuristic is used to compare the pairs that creates selective pressures for lexicons with low ambiguity, low synonymy, frequent meanings, and generality. The algorithm was tested on a noiseless corpus of sentences in English, Spanish, Turkish and Japanese, each sentence paired with an appropriate logical query. The authors show that, assuming one single meaning is possible for each sentence, their algorithm outperforms Siskind's. However, Siskind's algorithm is more general in that it handles noise and referential uncertainty. Also, unlike Siskind's incremental approach, Thompson and Mooney's algorithm batches, and they do not argue for psychological plausibility.

4.2 Learning from form-meaning pairs

In this section, we review a number of statistical models which receive pairs of form and meaning as input. The examined models pursue different goals, such as learning lexical constructions (Chang, 2004), learning verb meaning from image data (Dominey, 2003; Dominey and Inui, 2004), acquiring the syntactic and semantic features of verbs (Niyogi, 2002) and setting parameters of UG (Buttery, 2003, 2004). All these models process natural language text, but depending on the purpose of the model, each one processes a different type of meaning representation (e.g., labeled graphs, image data, feature vectors, semantic symbols).

The most similar work to ours is Chang's (2004). She presents a model for learning lexically specific multi-word constructions from annotated child-directed transcript data. The goal of the model is to learn associations between form relations (typically word order) and meaning relations (typically role-filler bindings) from input data, and to use them in language comprehension. The learning task is defined as finding the best grammar to fit the observed data given the space of possible grammars and a sequence of training examples consisting of an utterance paired with its context, where the space of possible grammars is defined by a unification-based formalism called Embodied Construction Grammar (ECG). In this framework, both form and meaning are represented as subgraphs of elements and relations among them, and lexical constructions involve a simple mapping between these two subgraphs. The prior knowledge embedded in the model consists of conceptual knowledge (an ontology of typed feature structures or schemata for people, objects, locations and actions) and lexical knowledge (a set of lexical constructions represented using the ECG formalism, linking simple forms to specific conceptual items). The model uses a construction analyzer that identifies the constructions responsible for a given utterance based on partial parsing techniques. Learning (updating the grammar) includes forming new structured maps to account for mappings present in the input but unexplained by the current grammar, and merging similar constructions into a more general or a larger one. The minimum description length (MDL) heuristic is used to evaluate the proposed constructions in terms of the size of the final grammar and the cost of the data given the grammar. The model learns only item-based constructions, and the only generalization learned over the input concerns the semantic constraints on the arguments (for example, THROW-BALL construction and THROW-BLOCK construction are merged into a general THROW-OBJECT construction), which is a minor achievement considering the complicated framework and algorithms the model employs. However, since this work is part of a larger project on language comprehension, it provides a testbed for applying the acquired knowledge in language tasks.

Another model for learning grammatical constructions by pairing form and meaning is presented by Dominey (2003) and Dominey and Inui (2004) in which a restricted kind of meaning is extracted from video images. The grammatical construction of the corresponding sentence is uniquely identified by the function words that appear in the sentence, and by the order of the words. Learned constructions are stored in a construction inventory, and are retrieved as needed using the function words (or closed class words) they contain as an index. Words in sentences and elements in the scene are encoded as two input vectors. Open-class words (nouns, verbs, adjectives and adverbs) and closed-class words (prepositions, determiners, etc.) are processed separately. Open class words are mapped onto their roles in the scene elements by a form-to-meaning mapping function, specific to each sentence type. This mapping is retrieved from the construction inventory using a construction index that encodes the closed-class words that characterize each sentence type. The model performs error-free role assignment for familiar events and word sets, and allows for limited generalization over new verbs with respect to assigning the right role to the scene elements. However, learning grammatical constructions is highly dependent on the presence of the closed class words, and testing the model is limited to distinctive constructions, i.e., those that contain different sets of closed class words or have different word order.

Niyogi (2002) presents a Bayesian model of syntactic and semantic bootstrapping which predicts the syntactic or semantic features of verbs based on a small set of scene-utterance pairs. Allowable syntactic alternations are represented by a hypothesis space, and the prior probability of all the hypotheses is assumed to be equal. Also, the model receives as prior knowledge the likelihood of observing evidence given a particular hypothesis. Possible verb meanings are modeled as a set of features, where each feature represents a predicate on one or more of the arguments of the verb, e.g., *moving*(x) or *contact*(x,y). The evidence from the scene is also represented as a feature vector. An idealized lexicon contains the word-concept mappings. Given a hypothesis space of possible verb concepts, the task of inducing a verb's meaning given a number of observations of scenes is to determine which of the possible concepts is the most likely. The model learns to predict missing features of verbs with more accuracy as it processes more input, and it is able to handle noise to some extent. However, this is not a real acquisition model, since the structure of the verb classes and their prior probabilities, as well as the probabilities of verbs showing particular features are all fixed. The model relies heavily on its embedded knowledge, including the structure of the hypothesis space, prior probabilities of each hypothesis, likelihood of each scene given a word concept, and perfect knowledge of the features of the complement. However, the model does not make any explicit claims about either the innateness or the learnability of this knowledge.

Buttery (2003, 2004) presents a nativist learning system, Categorial Grammar Learner (CGL), that employs a Bayesian parameter setting algorithm to acquire a lexicon. The syntactic properties of a word are contained in its lexical entry in the form of a syntactic category constructed from a finite set of primitive categories combined with two operators, `\` and `/` (forward and backward composition). The learning system is composed of three modules: a semantic learning module which learns the mapping between word tokens and semantic symbols; a syntax learning module which attempts to create valid parse trees for each utterance, and resolve unknowns in the syntactic category types; and a memory module which records the current state of the hypothesis space. This state consists of two hierarchies of parameters which must be set: categorial parameters, which determine whether a category is in use within the learner's current model of the input language; and word order parameters, which determine the underlying order in which constituents occur. Categories *S* (sentence), *SB* (subject) and *OB* (object) are primitive categories which are innate to the learner, and the syntactic categories of some group of words are determined prior to learning. The learner works based on the fundamental assumption that the semantic arity of a word is the same as its number of syntactic arguments, which is restrictive in some cases. The model also cannot handle those cases where some arguments are not syntactically realized (e.g., agent in *the hammer broke the window*). They tested their learner on a subset of CHILDES for which a unification based grammar description has been created, and achieved a recall of 68% with a precision of 98% for the semantic learner (the semantic learner is based on Siskind's (1996)). The syntactic learner is dependent upon the recall of the semantic learner; for those cases where a syntactic category was available for every word in the utterance, the syntactic learner achieved a precision of 82%.

4.3 Learning grammar from text

The models reviewed in this section exploit corpus-based learning algorithms to extract a grammar based on the distributional characteristics of the corpus. They use different machine learning techniques to learn the optimized grammar that accounts for the text they process as input. Although none of these models use any knowledge of semantics for grammar induction, they demonstrate the possibility of learning a large-scale grammar for a natural language from data.

MOSAIC (Model Of Syntax Acquisition In Children) (Jones et al., 2000; Gobet et al., 2004) is a computational model that learns from raw text, and produces utterances similar to what children produce. MOSAIC analyzes the distributional characteristics present in the input using two discrimination- and generalization-based learning mechanisms. The first mechanism grows an n-ary discrimination network consisting of nodes connected by directed test links, where nodes encode single words and test links encode the difference between the contents of the connected nodes. The second mechanism creates a new type of link, a *generative link*, between two nodes which share a certain percentage (set to 10%) of words surrounding them. The model generates output by traversing the network and outputting the contents of the visited links. MOSAIC was trained on a subset of CHILDES, and used to simulate a number of phenomena in language acquisition, including the verb island phenomenon (Jones et al., 2000) and the optional infinitive phenomenon in English (Gobet et al., 2004). In each case, the output of the model was compared to what children uttered. The use of generative links enables the model to demonstrate limited generalization abilities. However, the lack of semantic knowledge prevents the model from performing any meaningful generalization, and the generalized sentences are limited to the high-frequency terms.

ADIOS (Automatic DIstillation Of Structure) (Solan et al., 2004) is another model for learning a hierarchical representation of linguistic structures from a large-scale corpus of untagged natural language sentences. The model employs an unsupervised pattern acquisition algorithm which represents sentences as paths on a graph whose vertices are words. The algorithm detects significant patterns, or similarly structured sequences of words that recur in the corpus, by recursively performing context-sensitive statistical inference. Each pattern has an associated equivalence class, which is a set of alternative words that may fit into the slot in the pattern to construct a given path through the graph. Linguistic constructions are represented by trees composed of significant patterns and their associated equivalence classes. The collection of patterns learned this way from a corpus can be considered as a kind of empirically determined construction grammar, which enables the model to generalize to unseen text. The model was trained on a subset of CHILDES, and the extracted patterns were used to perform a grammaticality judgment test as done in English as a Second Language (ESL) classes. It achieved an accuracy of around 60%. However, since the model possesses no semantic knowledge, the ability to produce unseen text is limited to frequent terms that participate in the equality classes.

Clark (2001b,a) presents an algorithm using context distribution clustering (CDC) for the unsupervised induction of stochastic context-free grammars from tagged text. Sets of tag sequences are clustered together based on the contexts they appear in, where the context consists of the part of speech tag immediately preceding the sequence and the tag immediately following it. Thus sequences of tags end up in the same cluster when their context distributions are similar. Some of the clusters correspond to syntactic constituents, but others correspond to parts of constituents, and must be removed. The criterion proposed for this purpose is that there is high mutual information between the symbol occurring before the specified constituent and the symbol after. This technique is incorporated into a grammar induction algorithm using a minimum description length approach. Starting with an empty grammar and a single non-terminal, at each iteration all frequent strings are clustered and filtered according to the mutual information criterion. The cluster that gives the best immediate reduction in description length is selected, and a new non-terminal is added to the grammar with rules for each sequence in the cluster. The corpus is partially parsed based on the new grammar, and rules with the same left hand side are aggregated. The process is repeated until the algorithm stops producing plausible constituents. The algorithm was trained on the BNC, then the ATIS corpus was tagged and partially parsed using the acquired syntactic categories and grammar rules. The results were evaluated against the gold-standard ATIS parse using the PARSEVAL metrics, and compared to two other algorithms. The proposed algorithm is computationally expensive, and is sensitive to the order in which the rules are acquired, but nevertheless it is capable of extracting a large-scale context-free grammar from text in an unsupervised fashion. No lexical information is learned by the algorithm.

Onnis et al. (2002) use minimum description length to investigate the role of the simplicity principle in recovering from overgeneralization, i.e., cognitive systems seek the hypothesis that provides the briefest representation of the available data. They formulate the problem as choosing the candidate model of the right complexity to describe the corpus data, as determined by the simplicity principle. A corpus of sentences from an artificial language was used, where two syntactic categories noun and verb form evenly distributed two-word sentences of the form NV or VN, and a number of particular lexical combinations are disallowed. This approach to language acquisition does not focus on how learning occurs. Rather, they compare different hypotheses (grammars) for corpora of different size. For each corpus size, they choose the grammar that minimizes the sum of the grammar-encoding length and the data-encoding length. To test their approach, they compare data compression of corpora by two similar models: one posits a completely regular rule, whilst the other posits a regular rule and some exceptions to it. The simulations show that the first model gives a simpler encoding for smaller corpus size, but as the size of the corpus increases, that model requires a longer code. Although the goal of this work is to suggest the simplicity principle as a major learning mechanism in children, the simulations presented do not conform to cognitively plausible restrictions. The language under study is processed in batches, no semantic knowledge is used, and word frequencies are ignored.

4.4 Connectionist models

A number of connectionist models have been proposed in recent years that attempt to learn syntactic structure from sequences of words. The most influential approach of this kind is due to Elman (1990, 1991), who trained a simple recurrent network to predict the next input word, for sentences generated by a small context-free grammar (simple recurrent networks have an additional set of units called context units which provide for a limited amount of recursion). In Elman (1990), words were chosen from thirteen classes of nouns and verbs, and were encoded by a vector in which each word was represented by a different bit. This encoding scheme guaranteed that vectors reflected nothing about the form class or meaning of the words. The task given to the network was to learn to predict the order of successive words. The network was trained by a sequence of vectors forming an input sequence, where each word in the sequence was input, one at a time, in order. The network learned to represent words by categorizing them as nouns or verbs, with further subcategorization of nouns as animate/inanimate, human/non-human, etc. These representations were developed by the network and were not explicitly taught.

Elman (1991) extended the above-mentioned model by training a simple recurrent network on a task involving stimuli in which there are underlying hierarchical and recursive relationships. The input is similar to the previous task, but the resulting sentences have some additional properties: subject nouns agree with their verbs; verbs have different argument structures; recursive relative clauses (beginning with *who*) are permitted; and an end-of-sentence marker is introduced which can potentially occur anywhere in a string where a grammatical sentence might be terminated. An analysis of the hidden unit activation patterns shows that the network accomplishes the task by developing distributed representations which encode the relevant grammatical relations and hierarchical constituent structure. Although Elman's model shows very well the possibility of categorizing and learning abstract structures from strings of words, it cannot be considered a cognitive model, as the artificial language it learns is purely syntactic, whereas natural language learning is crucially an attempt to discover the relationship between meaning and linguistic form.

In an attempt to fill the gap between the learning of syntax and semantics in connectionist models, McClelland and Kawamoto (1986) present a connectionist model for assigning roles to constituents of sentences, simply by using the surface structure of the sentence as the input. Allen (1997) next presents a connectionist model for thematic role assignment which simulates the integration of syntactic, semantic and lexical information, including the number of arguments in the clause, the identity of the verb, the semantics of the verb, the presence and identity of any preposition, and the identity and order of arguments in the utterance. While simple localist representations are used for verbs and prepositions in the utterance, a distributed representation of nouns is used, to allow for generalization over semantic features. The position of arguments in the utterance is represented by distinct temporal patterns of presentation. The output array includes a set of semantic features for the verb, and a set of proto-role features for the arguments. The network was trained using a corpus consisting of the representations of the caretaker speech utterances containing the 110 most frequent verbs from CHILDES. Testing the network consisted of supplying both grammatical and ungrammatical novel utterances and measuring the activations of role units. The network shows the capacity to generalize and to distinguish the grammatical and ungrammatical utterances. The network also shows the ability to guess the semantic features of novel verbs. However, the model is confined to a very limited grammar and vocabulary which are not easily extendable. Function words (except for some prepositions) are ignored, and utterances with clausal arguments cannot be processed. The semantics of each verb is assumed to be constant in all situations, i.e., each verb is represented by a fixed semantic feature vector regardless of the context it is used in. Also, the knowledge of language learned by the model can be used only in a very limited way, and no production based on the acquired knowledge takes place in this model.

Allen's (1997) model was later extended by Allen and Seidenberg (1999) to propose a theory of grammaticality judgment. The model is composed of a comprehension layer which, given a sequence of words as input, computes their semantic representations, and a production layer which, given an input sequence of meanings, computes the appropriate words. The judgment process is modeled by querying the network for its version of an input sentence, and then measuring how far apart the input and output forms are. The semantics of each utterance is represented by sequences of individual word-level semantic representations, so no relationship or binding information is captured. The form of each utterance is also represented as a series of words presented over time.

Howell et al. (2001) present a neural network model for learning both lexical and grammatical aspects of language. Their model has two layers of input: the semantic features, and the words in the vocabulary. A semantic output layer is paired with the semantic input layer, and a linguistic auto-encoder and the linguistic predictor are paired with the linguistic input layer. The model memorizes both its linguistic and semantic inputs, and associates the two together, such that either one elicits the other through prediction. The authors show that although grammar exposure begins at the same time as lexical learning, grammar learning doesn't effectively take place until the lexical representations are solidified. Multiple runs of the network show that the tasks are learned in the expected order. A thorough evaluation of the work was not possible due to the lack of information provided in the paper.

Another connectionist model for learning language is presented by Desai (2002), where a neural network attempts to learn a miniature language given some sentences of this language and the corresponding scenes which the sentences describe. Sentences describe one or two objects and optionally an action, generated by a simple grammar. The sentences are presented to the network sequentially, one word per time step. The output of the network is the description of the corresponding scenes as a fixed-width vector, consisting of two slots for objects and one slot for the action, where each slot is consisted of 10-bit units. The network was trained on a subset of the sentences generated by the grammar, with different probabilities attached to each sentence type. The remaining sentences were used as test set, 96.7% of which were processed correctly after training. The network shows syntactic bootstrapping behaviour, since it can use the syntactic context to predict a component of the meaning of a novel word. Also, the network's learning of new syntactic frames is aided when the input contains familiar words. The authors argue that this can be described as semantic bootstrapping in a general sense. However, the semantic knowledge learned and used by the model is very limited (*size* feature for objects and *action* and *causality* features for the actions). The model only performs limited generalization for sentence comprehension, and the learned grammar is used in no other way such as sentence production or role assignment.

5 Summary

We reviewed evidence from psycholinguistic experimental data, showing that argument structure acquisition in children is initially item-based, and general argument structure constructions are learned later. The psycholinguistic theories that address this process do not offer a detailed account of how general knowledge is induced from the instances of input. By contrast, computational models can provide some insight about the underlying learning mechanisms, algorithms and structures exploited by human beings during the acquisition process. Based on what we examined in this paper, a comprehensive computational model for argument structure acquisition should:

- conform to the basic cognitive plausibility constraints, such as incrementality, and memory limitations;
- take the distributional characteristics of the
- use only positive evidence as input;
- exploit both syntactic and semantic aspects of input to learn the relationship between form and meaning;
- handle noisy data, including
- account for the item-based
- provide a mechanism to induce general constructions that pair form with meaning from the knowledge learned for
- explain how the acquired lexical and general knowledge are used for different language tasks such as production and comprehension;
- account for the errors that children make during the learning process, and the gradual recovery from these errors;
- be extendable, in the size of the vocabulary, and the complexity of the syntactic constructions and the semantic representation that it can cover.

A number of computational models have been studied in this report. These models are proposed for different purposes, and use a variety of learning algorithms, underlying representations, and theoretical backgrounds. However, at the time of writing this document, there is no computational model we are aware of that addresses all the above issues.

References

Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26:339–356.

Allen, J. (1997). Probabilistic constraints in acquisition. In Sorace, A., Heycock, C., and Shillcock, R., editors, *Proceedings of the GALA97 Conference on Language Acquisition, Edinburth*, pages 300–305.

Allen, J. and Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In MacWhinney, B., editor, *Emergence of Language*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bencini, G. M. L. and Goldberg, A. E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43:640–651.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni di semantica*, 3:5–66.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: are children helped by innate linking rules? *Linguistics*, 28:1253–1289.

Bowerman, M. (1996). Argument structure and learnability: is a solution in sight? In *Proceedings of the Berkeley Linguistics Society*, volume 22, pages 454–468.

Briscoe, T. (1999). The acquisition of grammar in an evolving population of language agents. *Machine Intelligence*, 16.

Briscoe, T. (2000). Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2).

Buttery, P. (2003). A computational model for first language acquisition. In *Proceedings of CLUK6, Edinburgh*, pages 1–8.

Buttery, P. (2004). A quantitative evaluation of naturalistic models of language acquisition: the efficiency of the Triggering Learning Algorithm compared to a Categorial

Grammar Learner. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 1–6.

Cartwright, T. A. and Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition*, 63:121–170.

Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 17–24.

Chomsky, N. (1986). *Knowledge of language*. Berlin: Praeger.

Christiansen, M. H. and Chater, N. (1999). Connectionist natural language processing: the state of the art. *Cognitive Science*, 23(4):417–437.

Clark, A. (2001a). Unsupervised induction of stochastic context free grammars with distributional clustering. In *Proceedings of CoNLL, Toulouse, France*, pages 105–112.

Clark, A. S. (2001b). *Unsupervised language acquisition: theory and practice*. PhD thesis, University of Sussex.

Demuth, K., Machobane, M., and Moloi, F. (2003). Rules and construction effects in learning the argument structure of verbs, in press. *Journal of Child Language*.

Demuth, K., Machobane, M., Moloi, F., and Odato, C. (2002). Rule learning and lexical frequency effects in learning verb-argument structure. In *Proceedings of the 26th Annual Boston University Conference on Language Development*, pages 142–153.

Desai, R. (2002). Bootstrapping in miniature language acquisition. In *Proceedings of the 4th International Conference on Cognitive Modelling*.

Dominey, P. F. (2003). Learning grammatical constructions in a miniature language from narrated video events. In *Proceedings of the 25nd Annual Conference of the Cognitive Science Society*.

Dominey, P. F. and Inui, T. (2004). A developmental model of syntax acquisition in the construction grammar framework with cross-linguistic validation in English and

Japanese. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 33–40.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Elman, J. L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225.

Fisher, C., Hall, G., Rakowitz, S., and Gleitman, L. (1994). When it is better to receive than to give: syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92(1):333–375.

Fodor, J. (1998). Unambiguous triggers. *Linguistic Inquiry*, 29(3):1–36.

Gibson, E. and Wexler, K. (1994). Triggers. *Linguistic Inquiry*, 25(3):404–454.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1:135–176.

Gobet, F., Freudenthal, D., and Pine, J. M. (2004). Modelling syntactic development in a cross-linguistic context. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 53–60.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. The University of Chicago Press.

Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In MacWhinney, B., editor, *Emergence of Language*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Howell, S. R., Becker, S., and Jankowicz, D. (2001). Modelling language acquisition: lexical grounding through perceptual features. In *Proceedings of the 2001 Developmental and Embodied Cognition Conference*.

Jones, G., Gobet, F., and Pine, J. M. (2000). A process model of children's early verb use. In *Proceedings of the 22th Annual Conference of the Cognitive Science Society*.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates, 2 edition.

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

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., and Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57(4, Serial No. 228).

McClelland, J. L. and Kawamoto, A. H. (1986). *Mechanisms of sentence processing: assigning roles to constituents of sentences*, pages 272–325. MIT Press Cambridge, MA, USA.

Miller, G. (1990). Wordnet: an on-line lexical database. *International Journal of Lexicography*, 17(3).

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*.

Onnis, L., Roberts, M., and Chater, N. (2002). Simplicity: a cure for overgeneralizations in language acquisition? In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, pages 720–725.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

Sakas, W. and Fodor, J. (2001). *The structural triggers learner*, chapter 5. Cambridge University Press, Cambridge, UK.

Siskind, J. M. (1994). Lexical acquisition in the presence of niose and homonymy. In *Proceedings of the American Association of Artificial Intelligence AAAI-94*, pages 760–766.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.

Solan, Z., Horn, D., Ruppin, F., and Edelman, S. (2004). Unsupervised context sensitive language acquisition from a large corpus. In Saul, L., editor, *Advances in Neural Information Processing*, volume 16. Cambridge, MA, MIT Press.

Thompson, C. A. and Mooney, R. J. (1999). Automatic construction of semantic lexicons for learning natural language interfaces. In *Proceedings of the 17th National Conference on Artificial Intelligence*, Orlando, FL.

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition*, 74:209–253.

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

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomasello, M. and Abbot-Smith, K. (2002). A tale of two theories: response to Fisher. *Cognition*, 83:207–214.