# Yale University
# Department of Computer Science

A Model of Semantics and Corrections
in Language Learning

Dana Angluin      Leonor Becerra-Bonache

# A Model of Semantics and Corrections in Language Learning

Dana Angluin[*]        Leonor Becerra-Bonache[†]

## Abstract

We present a computational model of language learning via a sequence of interactions between a teacher and a learner. The utterances of the teacher and learner refer to shared situations, and the learner uses cross-situational correspondences to learn to comprehend the teacher's utterances and produce appropriate utterances of its own. We show that in this model the teacher and learner come to be able to understand each other's meanings. Moreover, the teacher is able to produce meaning-preserving corrections of the learner's utterances, and the learner is able to detect them.

We test our model with limited sublanguages of several natural languages in a common domain of situations. The results show that learning to a high level of performance occurs after a reasonable number of interactions. Moreover, even if the learner does not treat corrections specially, in several cases a high level of performance is achieved significantly sooner by a learner interacting with a correcting teacher than by a learner interacting with a non-correcting teacher. Demonstrating the benefit of semantics to the learner, we compare the number of interactions to reach a high level of performance in our system with the number of similarly generated utterances (with no semantics) required by the ALERGIA algorithm to achieve the same level of performance.

We also define and analyze a simplified model of a probabilistic process of collecting corrections to help understand the possibilities and limitations of corrections in our setting.

## 1   Introduction

We investigate aspects of the roles of semantics and corrections in the process of learning to understand and speak a natural language. A child learning his or her native language typically does so while interacting with other people who are using the language to communicate in shared situations. The correspondence between situations and utterances seems likely to be a very important source of information for the language learner. Once a child begins to produce his or her own utterances, other people's responses to them (or lack thereof) are another source of information about the language. When the child's utterances fall short of adult-level competence, sometimes the other person in the conversation will repeat the child's utterance in a more correct form. A number of studies have focused on the phenomenon of such corrections and questions of their frequency in

child-directed speech and whether children can and do make use of them; some of these studies are discussed in the next section.

In this paper we construct a computational model with a learner and a teacher who interact in a sequence of shared situations, extending our previous work on semantics and comprehension [2,3] to model production and correction. In each situation the teacher and learner interact as follows. First the learner uses what it has learned about the language to (attempt to) generate an utterance appropriate to the situation. The teacher then analyzes the correctness of the learner's utterance and either generates an utterance intended as a correction of the learner's utterance, or generates another utterance of its own appropriate to the situation. Finally, the learner uses information given by its own utterance, the teacher's utterance and the situation to update its knowledge of the language. At the conclusion of this interaction, a new interaction is begun with the next situation in the sequence.

Both the learner and the teacher engage in comprehension and production of utterances which are intended to be appropriate to their shared situation. This setting allows us to study several questions: whether access to the semantics of the shared situations facilitates language acquisition by the learner, whether the teacher can offer meaningful corrections to the learner, whether the learner can detect intended corrections by the teacher, and whether the presence of corrections by the teacher has an effect on language acquisition by the learner. For our model, the answer to each of these questions is yes, and while the model is in many respects artificial and simplified, we believe it sheds new light on these issues.

## 2   Related Work

In this section we review work on computational models of language learning incorporating semantics, and work in linguistics related to the issues of positive and negative data and corrections in natural language acquisition.

### 2.1   Computational models of language learning

The study and development of computational models of language learning is of great interest not only to better understand the process of language acquisition, but also for practical applications of language learning by machines; the advantages of having a machine able to understand and speak a natural language would be innumerable. However, the complexity of these tasks has led researchers to focus on individual sub-problems. As a result most work has focused only on language comprehension, and on just a single phenomenon (e.g., single word learning [26] or syntactic category acquisition [24].) The problem of focusing on just one phenomenon is that we can have a model that is good at performing a task but cannot perform others tasks (for example, it can be good at lexical acquisition but not at syntactic acquisition.)

Tellier [33] gives a very interesting theoretical account of the possible relationship between semantics and syntactic learning, and suggests that "the acquisition of a conceptual representation of the world is necessary *before* the acquisition of the syntax of a natural language can start." Our system reflects a similar assumption that the learner embarking on language learning has a considerable stock of semantic knowledge.

Some of the more empirical works that are focused on comprehension and take into account semantics are those of Siskind [31,32], Marcken [10], Regier [25] Bailey [4], Roy and Pentland [28]

2

and Yu and Ballard [37]. The systems developed by Siskind [31] and Marcken [10] learn word semantics in simulation; it is assumed that the learner already has concepts for each of the words to be learned, and word learning becomes a problem of matching words to concepts. The representation used by these systems is based on compositional semantics. Moreover, Siskind investigated the use of cross-situational analysis to model lexical acquisition. The systems introduced by Regier [25] and Bailey [4] use more realistic simulated data; they focus on learning particular words in isolation, rather than in a linguistic context. And finally, systems using embodied sensors, such as that of Roy and Pentland [28] and Yu and Ballard [37], have focused on learning definitions for concrete nouns or physical verbs. The learner does not know anything about the language, except its phonemes. These systems can produce associations between phoneme sequences and visual stimuli, but they cannot produce grammatical utterances from what they have learned.

One of the most closely related research efforts is that of Kevin Gold and his collaborators [13–15]. They introduced a word learning system called TWIG (Transportable Word Intension Generator) that allows a robot to learn compositional meanings for new words. An important feature of TWIG is the use of decision trees to build word definitions. Unlike other approaches, TWIG learns tree structures for storing word semantics in an unsupervised manner. Moreover, it uses words learned in complete grammatical sentences for production, comprehension, or referent inference. Our model also accommodates comprehension and production tasks, but we use a different representation of word meaning, that is not based on Frege's principle of compositionality; this allows us to assign a correct meaning to a sentence even if the sentence is not grammatically correct. We also use decision trees for learner production, not to provide the definition of a word but to help the learner to decide the choice of a phrase in a given position. (Decision trees of this kind were also used in an earlier version of TWIG.) Unlike in TWIG, our learner has no initial knowledge of words or grammar, there is no restriction on the number of new words in a sentence given to the learner, and our learner is allowed to interact with the teacher and possibly receive corrections.

Our work is also related to that of Jack [19]. He introduced a computational model that produces behavior comparable to the stages of children's linguistic development. The model takes syllable-segmented input, learns to associate words with meanings under referential uncertainty, discovers compositional combinations of words, learns to make appropriate use of word ordering, and derives grammar rules and classes describing a language fragment. Our model has clear connections with Jack's model, but our learner is quite different from his learner: the input given to our learner can contain sentences that describe the situation only partially; our learner interacts with the teacher and can be corrected; our learner is not able to produce any sentences after receiving just one sentence from the teacher; and our learner can produce sentences that have a telegraphic character.

A common feature of many of these approaches is the use of a miniature language environment. A task known as *Miniature Language Acquisition* was formalized by Feldman et al. [18]. The task consists of learning a subset of a natural language from sentence-picture pairs that involve geometric shapes with different properties (color, size and position.) Although this task is not as complex as those faced by children, it involves enough complexity to be compared to many real-word tasks. We use a simplified version of Feldman's task.

Other works that have also provided inspiration for our model are by Collins Hill [16] and Shaerlaekens [30]. They model the language of 2-year old children, and point out the relevance of semantics for the construction of the initial grammar of the child. Collins Hill introduced a repetition and response model in which the model produces responses to or repetitions of adult

3

input sentences. This model fits the data collected from a single 2-year old girl, Claire. An interesting feature of this model is the use of a grammar composed of templates, that is, patterns for understanding and producing speech. Based on evidence provided by the study of two sets of Dutch-speaking triplets, Schaerlaekens proposed a semantic relations model focused on the 2-word stage. Schaerlaekens considers that 2-word sentences are semantically interpretable and have their own syntax. She also points out the importance of the context to analyze and interpret 2-word sentences. Using the semantic relations model, Schaerlaekens constructs a grammar that contains a number of possible sentence patterns (each expressing one semantic relation) found in children's 2-word sentences. Using this methodology, she constructs a descriptive grammar for the first 200 2-word sentences of each of the six children in the study.

## 2.2  Positive versus negative evidence in language acquisition

Formal models of language acquisition have mainly focused on learning from positive data, that is, utterances that are grammatically correct. But a question that remains open is: Do children receive negative data and can they make use of it?

There have been three main proposals related to this question: i) Children do not receive negative information and rely only on innate information; ii) Children receive negative information in the form of different reply-types given in response to ungrammatical versus grammatical child utterances; iii) Children receive negative information in the form of reformulations (that is, utterances adults use in checking up on what children intended to say), and they not only can detect them but also make use of them. The issue of whether or not children receive negative evidence depends in part on how one defines this concept. Therefore, it seems important to define negative evidence exactly. Next we review these three proposals.

Chomsky's poverty of stimulus argument has been used to support the idea of human innate linguistic capacity. It is claimed that there are principles of grammar that cannot be learned from positive data only, and negative evidence is not available to children. Hence, since children do not have enough evidence to induce the grammar of their native language, the additional knowledge language learners need is provided by some form of innate linguistic capacity.

E. M. Gold's negative results in the framework of formal language learning have also been used to support this position. Gold proved that superfinite languages are not learnable from positive data only, which implies than none of the language classes defined by Chomsky to model natural language is learnable from positive data [12]. However, he suggests several hypotheses about how children overcome this theoretical hurdle. The first hypothesis is that the class of possible natural languages is much smaller than we would expect from our current models of syntax. The second hypothesis is that the child receives negative information by being corrected in a way we do not recognize. And a third possibility is that there is an a priori restriction on the class of texts that can occur, such as a restriction on the order of text presentation. Theoretical works have shown that the first hypothesis can lead to successful learning (e.g., results on learning restricted classes of formal languages using positive data by Angluin [1], Sakakibara [29] and Kanazawa [20].) In linguistics, it is also generally assumed that the first hypothesis holds. We suggest that it is also worth exploring the second hypothesis pointed out by Gold.

Brown and Hanlon [5] studied negative evidence understood as explicit approvals or disapprovals of a child's utterance (e.g.,"That's right" or "That's wrong.") They showed that there is no dependence between these kinds of answers and the grammaticality of children's utterances. These results were taken as showing that children do not receive negative data. But do these results really

show this? It seems evident that parents rarely address their children in that way. During the first stages of language acquisition children make a lot of errors, and parents are not constantly telling them that their sentences are wrong; rather the important thing is that they can communicate with each other. However, it is worth studying whether other sources of negative evidence are provided to children. Is this the only form of negative data? Do adults correct children in a different way?

Some researchers have studied other kinds of negative data based on reply-types (e.g., Hirsh-Pasek et al. [17], Demetras et al. [11] and Morgan and Travis [23].) These studies argue that parents provide negative evidence to their children by using different types of reply to grammatical versus ungrammatical sentences. Marcus analyzed such studies and concluded that there is no evidence that this kind of feedback (he called it noisy feedback) is required for language learning, or even that it exists [22]. He argued for the weakness, inconsistency and inherently artificial nature of this kind of feedback. Moreover, he suggested that even if such feedback exists, a child would learn which forms are erroneous only after complex statistical comparisons. Therefore, he concluded that internal mechanisms are necessary to explain how children recover from errors in language acquisition.

Since the publication of the work of Marcus, the consensus seemed to be that children do not have access to negative data. However, a study carried out by Chouinard and Clark shows that this conclusion may be wrong [7]. First, they point out that the reply-type approach does not consider whether the reply itself also contains corrective information, and consequently, replies that are corrective are erroneously grouped with those that are not. Moreover, if we consider only reply-types, they may not help to identify the error made. Hence, Chouinard and Clark propose another view of negative evidence that builds on Clark's principle of contrast [8, 9]. Parents often check up on a child's erroneous utterances, to make sure they have understood them. They do this by reformulating what they think the child intended to express. Hence, the child's utterance and the adult's reformulation have the same meaning, but different forms. Because children attend to contrasts in form, any change in form that does not mark a different meaning will signal to children that they may have produced an utterance that is not acceptable in the target language. In this way, reformulations identify the locus of any error, and hence the existence of an error. Chouinard and Clark analyze longitudinal data from five children between two and four years old, and show that adults reformulate erroneous child utterances often enough to help learning. Moreover, these results show that children not only detect differences between their own utterance and the adult reformulation, but that they make use of that information.

Our model is inspired also by Chouinard and Clark's results. Corrections (in form of reformulations) have a semantic component that has not been taken into account in previous studies. Hence, we propose a new computational model of language learning that gives an account of meaning-preserving corrections, and in which we can address questions such as: What are the effects of corrections on learning syntax? Can corrections facilitate the language learning process? Can semantic information simplify the learning problem?

## 3 The Model

We describe the components of our model, and give examples drawn from the primary domain we have used to guide the development of the model.

## 3.1 Situation, meanings and utterances

A **situation** is composed of some objects and some of their properties and relations. These are not intended to be an exhaustive description of the state of the world, but to pick out some aspects of it that are of joint interest to the teacher and the learner. We assume that these objects, properties and relations are recognizable to both the learner and teacher from the outset. A situation is represented as a set of ground atoms over some constants (denoting objects) and predicates (giving properties of the objects and relations between them.) For example, a situation $s_1$ consisting of a big purple circle to the left of a big red star is represented by the following set of ground atoms:

$$s_1 = \{bi1(t_1), pu1(t_1), ci1(t_1), le2(t_1, t_2), bi1(t_2), re1(t_2), st1(t_2)\}.$$

Here the two objects are represented by the constants $t_1$ and $t_2$, and the unary predicates are big (*bi1*), purple (*pu1*), circle (*ci1*), red (*re1*) and star (*st1*), and the sole binary predicate is the relation of one object being to the left of another (*le2*). The teacher and learner each have access to this representation of the situation.

Formally, we have a finite set $P$ of **predicate symbols**, each of a specific arity (number of arguments), where the arity is appended to the name of the predicate. We also have a countable set of **constant symbols** $t_1, t_2, \ldots$, which are used to represent distinct objects. A **ground atom** is an expression formed by applying a predicate symbol to the correct number of constant symbols as arguments, for example, *re1(t₁)* or *ab2(t₂, t₁)*.

We also have a countable number of **variables** $x_1, x_2, \ldots$. A **variable atom** is an expression formed by applying a predicate symbol to the correct number of variables as arguments, for example, *re1($x_2$)* or *le2($x_4, x_3$)*. A **meaning** is a finite sequence of variable atoms. Note that the atoms do not contain constants, and the order in which they appear is significant. Examples of meanings are $m_1 = (st1(x_1))$, $m_2 = (st1(x_1), le2(x_2, x_1), pu1(x_2), ci1(x_2))$, $m_3 = (bi1(x_1))$ and $m_4 = (pu1(x_1), ci1(x_2))$.

A meaning is **supported** in a situation if there exists a **support witness**, that is, a mapping of its variables to *distinct* objects in the situation such that the image under the mapping of each atom in the meaning appears in the situation. If a meaning is supported in a situation by a unique support witness then it is **denoting** in the situation.

The meaning $m_1$ is supported in the situation $s_1$ via the witness $x_1 \rightarrow t_2$, the meaning $m_2$ is supported in the situation $s_1$ via the witness $x_1 \rightarrow t_2, x_2 \rightarrow t_1$, and the meaning $m_3$ is supported in the situation $s_1$ by either the witness $x_1 \rightarrow t_1$ or the witness $x_1 \rightarrow t_2$. The meaning $m_4$ is not supported in the situation $s_1$ because any support witness must map the variables $x_1$ and $x_2$ to distinct objects. Meanings $m_1$ and $m_2$ are denoting in the situation $s_1$, but meaning $m_3$ is not; intuitively, the predicate *bi1* by itself could refer to either object. We assume that both the teacher and learner can determine whether a meaning is denoting in a situation.

We also have a finite alphabet $W$ of words. An **utterance** is a finite sequence of words, for example, *the star* or *the star to the right of the purple circle* or *star of circle small the green*. The **target language** is the set of utterances the teacher may produce in some situation; in our example, this includes utterances like *the star* or *the star to the right of the purple circle* but not *star of circle small the green*. We assume each utterance in the target language is assigned a unique meaning; in our examples, a finite state transducer is used both to recognize utterances in the target language and to assign them meanings.

An utterance is **denoting** in a situation if the meaning assigned to utterance is denoting in the situation. Intuitively, an utterance is denoting if it uniquely picks out the objects it refers to in a

situation. Thus, in this model, an utterance is "appropriate" to a situation if it is denoting in that situation. For example, if the utterance *the star* is assigned the meaning $m_1$ then it is denoting in the situation $s_1$ and if the utterance *the star to the right of the purple circle* is assigned the meaning $m_2$, it is also denoting in the situation $s_1$. If the situation contained two stars, then the utterance *the star* would not be denoting. Note that a denoting utterance does not have to be "minimal" – it may specify more information than necessary to pick out one object.

In our model the goal of the learner is to be able to produce every appropriate utterance in any given situation. That is, given a situation, the learner should eventually produce only correct denoting utterances for that situation, and should be able to produce all of them. In fact, our model is probabilistic, and what we require is that the probability of learner errors be reduced to very low levels.

## 3.2 The target language and meaning transducers

We would like a method of specifying the linguistic competence of the teacher, including the target language of utterances and their meanings. This representation should enable production by the teacher, that is, given a situation, the teacher should be able to choose an utterance in the target language that is denoting for the situation. For the model developed in this paper, we represent this competence by a finite state transducer that both recognizes the utterances in the target language and translates each correct utterance to its meaning.

Recall that $W$ is the finite set of words and $P$ is a finite set of predicates. Let $A$ denote the set of all variable atoms over $P$. We define a **meaning transducer** $M$ with input symbols $W$ and output symbols $A$ as follows. $M$ has a finite set $Q$ of states, an initial state $q_0 \in Q$, a finite set $F \subseteq Q$ of final states, a deterministic transition function $\delta$ mapping $Q \times W$ to $Q$, and an output function $\gamma$ mapping $Q \times W$ to $A \cup \{\varepsilon\}$, where $\varepsilon$ denotes the empty sequence.

The transition function $\delta$ is extended to define $\delta(q, u)$ to be the state reached from $q$ following the transitions specified by the utterance $u$. The **language** of $M$, denoted $L(M)$ is the set of all utterances $u \in W^*$ such that $\delta(q_0, u) \in F$. A state $q \in Q$ is **live** if there exists an utterance $u$ such that $\delta(q, u) \in F$, and **dead** otherwise. For each utterance $u$, we define the **output** of $M$, denoted $M(u)$, to be the finite sequence of non-empty outputs produced by starting at state $q_0$ and following the transitions specified by $u$, that is, $M(u)$ is the **meaning** of $u$.

As an illustration, we describe a limited sublanguage of Spanish involving geometric shapes and their properties and relative locations. $W$ contains the words *el*, *la*, *circulo*, *cuadrado*, *triangulo*, *rojo*, *azul*, *verde*, *a*, *izquierda*, *derecha*, *encima*, and *del*. $P$ contains the predicate symbols *ci1*, *sq1*, *tr1*, *bi1*, *re1*, *bl1*, *gr1*, referring to the properties of being a circle, a square, a triangle, big, red, blue, green, and also *le2* and *ab2*, referring to the relations of one object being to the left of or above another object. Note that in this example there is a predicate for big but no word for big in the vocabulary. Also, there are words in $W$ for the relations left (*izquierda*), right (*derecha*) and above (*encima*), but the word for below (*debajo*) is omitted from $W$ in this example sublanguage.

We define a meaning transducer $M_1$ for the limited sublanguage as follows. The states are 0 through 10; 0 is the initial state and the final states are $\{2, 3, 8, 9\}$. The transition function is given in Table 1, and the automaton is pictured in Figure 1. Unspecified transitions go to the non-final dead state, 10.

The language of this transducer is finite and contains 444 utterances each with a distinct meaning. Examples are *el triangulo rojo* and *el circulo a la derecha del triangulo azul*, which have meanings of $(tr1(x_1), re1(x_1))$ and $(ci1(x_1), le2(x_2, x_1), tr1(x_2), bl1(x_2))$, respectively.

| state | word | next state | output |
|---|---|---|---|
| 0 | *el* | 1 | $\varepsilon$ |
| 1 | *circulo* | 2 | *ci1* $(x_1)$ |
| 1 | *cuadrado* | 2 | *sq1* $(x_1)$ |
| 1 | *triangulo* | 2 | *tr1* $(x_1)$ |
| 2 | *rojo* | 3 | *re1* $(x_1)$ |
| 2 | *azul* | 3 | *bl1* $(x_1)$ |
| 2 | *verde* | 3 | *gr1* $(x_1)$ |
| 2 | *a* | 4 | $\varepsilon$ |
| 3 | *a* | 4 | $\varepsilon$ |
| 4 | *la* | 5 | $\varepsilon$ |
| 5 | *izquierda* | 6 | *le2* $(x_1, x_2)$ |
| 5 | *derecha* | 6 | *le2* $(x_2, x_1)$ |
| 2 | *encima* | 6 | *ab2* $(x_1, x_2)$ |
| 3 | *encima* | 6 | *ab2* $(x_1, x_2)$ |
| 6 | *del* | 7 | $\varepsilon$ |
| 7 | *circulo* | 8 | *ci1* $(x_2)$ |
| 7 | *cuadrado* | 8 | *sq1* $(x_2)$ |
| 7 | *triangulo* | 8 | *tr1* $(x_2)$ |
| 8 | *rojo* | 9 | *re1* $(x_2)$ |
| 8 | *azul* | 9 | *bl1* $(x_2)$ |
| 8 | *verde* | 9 | *gr1* $(x_2)$ |

Table 1: Transitions and outputs of the transducer $M_1$.

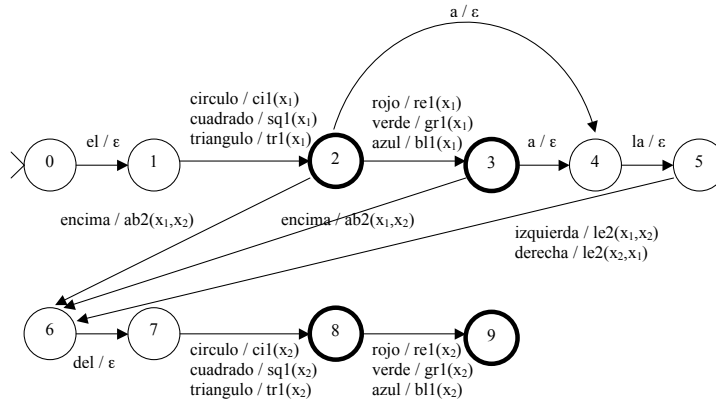

Figure 1: Meaning transducer $M_1$.

### 3.3   The learning task

The learning task is characterized by a meaning transducer, a process that produces situations, and the behavior of the teacher. The goal of the learner is to learn a grammar for the language that will enable it to produce all and only the denoting utterances for any given situation. [1] The learner gathers information about the language by engaging in a sequence of interactions with the teacher, where each interaction is related to a new situation.

Initially the teacher and learner know all of the predicates $P$ that will occur in situations, and are able to determine whether a given meaning is denoting in a given situation. A situation is a collection of objects, properties and relations, given to the learner and teacher as a set of ground atoms. Thus the learner and teacher share the same description of the situation.

The learner and teacher both also know a shared set of **categories** that classify a subset of the predicates into similarity groups. The categories facilitate generalization by the learner, and are used by the teacher in analyzing incorrect utterances of the learner. In our geometric shape examples, shape predicates are one category, color predicates are another, and size predicates are a third category, but no category contains the positional relations.

Initially the teacher also has the meaning transducer for the target language, which it uses both to produce utterances appropriate to the given situation, and to analyze the utterances of the learner. Initially the learner has no language-specific knowledge. The interactions of the learner and teacher are described in detail in the next section.

## 4   The Interaction of Learner and Teacher

Figure 2 summarizes one interaction between learner and teacher. In this section we describe the algorithms used by the learner and teacher to carry out the steps of this process.

### 4.1   Beginning to learn words: the co-occurrence graph

In each interaction, the learner is presented with a new situation, and attempts to produce an utterance appropriate to that situation. Because the learner initially has no language-specific knowledge, for the first few interactions the learner produces no utterance. When the learner produces no utterance, there is nothing for the teacher to respond to, so the teacher produces a randomly drawn denoting utterance for the situation, and the learner receives that utterance and uses it to learn more about the target language and its semantics.

In processing the teacher's utterance, the learner records the words that occur in it and the predicates that occur in the corresponding situation in a **co-occurrence graph** $G$. The nodes of the graph correspond to words and predicate symbols, and there is an undirected edge between every pair of nodes, corresponding to two words, two predicate symbols, or a word and a predicate symbol. Each node $u$ has an occurrence count, $c(u)$, recording the number of utterances or situations its corresponding word or predicate symbol has occurred in. Each edge $(u, v)$ also has an occurrence count, $c(u, v)$, recording the number of utterance/situation pairs in which the word or predicate symbol at the two endpoints have occurred together.

From $G$ the learner derives another graph, the **implication graph** $H$, which is a directed graph with the same set of nodes as $G$ and edges defined as follows. For each ordered pair of

---

[1] We refer to the learner's representation as a grammar, although it does not take the form of a classical grammar from formal languages.

Interaction of Learner and Teacher

1. A new situation is generated and presented to the learner and teacher.

2. The learner uses its current grammar to attempt to produce an utterance appropriate to the situation.

3. The teacher analyzes the learner's utterance (if any) in the context of the situation. The learner's utterance may be correct, have an error in form or an error in meaning, or be uninterpretable.

4. If the learner's utterance has an error in form or an error in meaning, the teacher decides randomly (using the correction probability) whether to correct the learner.

5. The teacher produces an utterance appropriate to the situation: either a correction of the learner's utterance or a randomly drawn denoting utterance.

6. The learner analyzes the teacher's utterance and updates its current grammar for the language as appropriate.

Figure 2: Summary of one interaction in the learning cycle.

nodes $(u, v)$, define $p(u, v) = c(u, v)/c(u)$. Intuitively, $p(u, v)$ may be thought of as a conditional probability: given that $u$ occurs in an utterance/situation pair, $p(u, v)$ is the probability that $v$ also occurs. If $p(u, v)$ is close to 1, then nearly every time $u$ occurs, $v$ also occurs. Because we want to be able to tolerate errors at some rate, there is a noise threshold $\theta$ (currently 0.95) and we say that $u$ **approximately implies** $v$ if $p(u, v)$ is not less than $\theta$. The directed edges $(u, v)$ of the implication graph are those pairs $(u, v)$ such that $u$ approximately implies $v$. The implication graph is recomputed after each situation/utterance pair.

If the teacher produces utterances using the meaning transducer $M_1$, after the learner receives a number of situation/utterance pairs, we expect that the word *cuadrado* will approximately imply the predicate symbol *sq1* and will not approximately imply any other predicate symbol, thus giving part of the meaning of the word *cuadrado*. Similarly, after a sufficient number of situation/utterance pairs, we expect that the word *rojo* will approximately imply the predicate symbol *re1* and the words *izquierda* and *derecha* will approximately imply the predicate symbol *le2*. Because the word *el* occurs in every situation/utterance pair, we expect that it will not approximately imply any predicate symbol. Thus, to a first approximation, it seems that the predicate symbol(s) (if any) approximately implied by a word should give part of the meaning of that word, after enough situation/utterance pairs.

As an example of some of the complications that arise in using this idea, consider the word *a*, which in this setting only occurs in the phrases *a la izquierda* and *a la derecha*. After sufficiently many situation/utterance pairs, the word *a* will also approximately imply the predicate symbol *le2* and will not imply any other predicate symbol, and similarly for the word *la*. It might be argued that this is an accidental consequence of the fact that we deal with a very limited sample of the language, which is true. However, children must deal robustly with a limited (but expanding) sample of the language they are learning, so it seems important to be able to cope in some way with such artifacts.

If we think about the relationships between the words *a*, *la*, *izquierda* and *derecha* in our setting, *izquierda* and *derecha* imply the presence of *a* and *la*, and the words *a* and *la* always co-occur. In some sense, the "proximate cause" of presence of the predicate symbol *le2* in the situation is the co-occurring pair of words *a* and *la*. This notion of "proximate cause" is captured graph-theoretically by taking a **transitive reduction** $H^r$ of the implication graph $H$ as follows. First, edges from predicates to words are removed. Then, each strongly connected component of the resulting graph is contracted to a vertex, the transitive reduction of the resulting acyclic graph is computed by removing all edges $(u, v)$ such that there is a directed path of length at least two from $u$ to $v$, and the strongly connected components are expanded again.

In our example the effect of this operation is to remove the implication edges from *izquierda* and *derecha* to the predicate symbol *le2*, while retaining the implication edges from *a* and *la* to the predicate symbol *le2*. In an expanded language sample in which *a* and *la* were used in situations not involving the predicate symbol *le2*, this artifact would disappear, and there would be implication edges from *izquierda* and *derecha* to the predicate symbol *le2*.

So, to summarize, the learner uses the situation and the teacher's utterance to update the co-occurrence graph, the implication graph, and the transitively reduced implication graph. The information contained in these structures is used by the learner in its attempts to produce its own utterances and to comprehend further utterances of the teacher.

## 4.2 Comprehension by the learner

Given a situation and an utterance by the teacher, the learner uses the transitively reduced implication graph $H^r$ in an attempt to determine the meaning of the teacher's utterance as follows. For each successive word of the teacher's utterance, the learner finds the list of all predicate symbols such that there is an edge from the word to the predicate symbol in $H^r$. Some of the predicate symbols may then be removed from each word's list as follows.

A **background predicate** is one whose predicate symbol has occurred in nearly all situations, that is, in a fraction of situations greater than the noise threshold $\theta$. Such a predicate symbol is approximately implied by every word, and is removed from the list of each word in the utterance. Also removed from each list is any predicate symbol whose arity is greater than the minimum arity of any predicate symbol on the list; thus, unary predicates are preferred over binary predicates if both are possible meanings. Finally, if two (or more) different words in the utterance are in the same strong component of $H^r$, all the predicate symbols are removed from the list of each of the equivalent words except for the one that occurs rightmost in the utterance. The result of this process is a list of lists of predicate symbols, one for each word in the teacher's utterance.

For example, early in one run of learning with $M_1$, the learner translated the teacher's utterance *el cuadrado rojo* into the following sequence of lists of predicate symbols:

$$((), (gr1, sq1), (re1)).$$

Evidently the data at that point were not sufficient to rule out the predicate symbol *gr1* as a possible meaning for *cuadrado*.

The learner forms a set of sequences of predicate symbols by taking (in order) one predicate from each non-empty list of predicate symbols in all possible ways. For the preceding example, that would produce two possible sequences of predicate symbols, namely, $(gr1, re1)$ and $(sq1, re1)$.

The resulting set $S$ of sequences of predicate symbols is then compared with the situation to try to determine the teacher's meaning. In particular, the learner computes a variable-normalized rep-

resentative of every meaning that is both supported by the situation and is such that its sequence of predicate symbols is in $S$ – these become the learner's guess of the **possible meanings** of the teacher's utterance in the given situation. To continue the preceding example, in the interaction in question, the situation consisted of a big red square above a big blue circle. Thus, the meaning $(sq1(x_1), re1(x_1))$ is both supported by the situation and has one of the two predicate sequences in $S$, namely, $(sq1, re1)$. This is the only possible meaning that the learner guesses for this utterance because there is no supported meaning with the predicate sequence $(gr1, re1)$, and no other inequivalent supported meaning with the predicate sequence $(sq1, re1)$.

If the learner finds exactly one possible meaning for the teacher's utterance, the learner takes this to be "the" meaning intended by the teacher, an assumption that may or may not be correct. In this case, the learner takes the unique meaning as the basis of a possible general form for meanings in the target language. Specifically, the learner uses its prior knowledge of the categories of the predicate symbols to generalize the unique meaning to a **general form**, by replacing each predicate symbol by its generalization (if any) in the set of categories. In our example, the category of $sq1$ is $shape1$ and the category of $re1$ is $color1$, so the learner generalizes the unique possible meaning $(sq1(x_1), re1(x_1))$ to the general form $(shape1(x_1), color1(x_1))$.

The general form is added to the (initially empty) set of possible general forms of meanings in the target language. The general forms acquired by the learner are the basis of the learner's own attempts to produce utterances relevant to a given situation. In particular, at any point, the current production grammar of the learner is represented by the list of general forms it has acquired, together with rules (not discussed yet) for how they may be instantiated.

To facilitate the "aging out" of incorrect general forms that are acquired early in the learner's experience, the learner records the interaction number of the most recent teacher utterance with a unique possible meaning that matched the general form. These values are used in the learner's production process in such a way that general forms that are not repeatedly matched to teacher utterances become less and less likely to be used by the learner.

If the learner finds no possible meanings, or more than one, for the teacher's utterance, the learner does not attempt to update the information in its set of general forms. In this model we do not attempt to quantify the learner's comprehension, preferring to measure the learner's progress by its productions. The comprehension process of the learner could be refined in various ways; for example, if there are several possible meanings and only one is compatible with the learner's current grammar, that one could be taken to be "the" teacher's meaning. We have not pursued this direction here.

## 4.3 Production by the learner

The general forms acquired by the learner are the basis on which it attempts to produce utterances appropriate to situations. Recall that each general form is a sequence of variable atoms, where the predicates may be from $P$ or may be category symbols. Each general form is a template denoting a set of possible meanings. That is, all the meanings obtained by substituting a corresponding predicate from $P$ for each category symbol is the set of meanings generated by that general form. For example, by substituting one of the predicates $sq1$, $ci1$ or $tr1$ for $shape1$ and one of the predicates $bl1$, $gr1$, or $re1$ for $color1$ in the general form $(shape1(x_1), color1(x_1))$, we get one of nine possible meanings, for example, $(ci1(x_1), bl1(x_1))$. The learner's categories are the basis it uses to generalize from the single meaning $(sq1(x_1), re1(x_2))$ to this set of nine possible meanings.

In attempting to produce an utterance appropriate to the current situation, the learner finds all

the meanings generated by its general forms using predicates from the current situation, and tests each meaning to see if it is denoting in the current situation, producing a set of possible denoting meanings for this situation. If the set is empty, the learner produces no utterance. Otherwise, it attempts to translate each denoting meaning into an utterance with that meaning, producing a set of possible denoting utterances; this process is described in the next section.

To produce one utterance for the current situation, the learner selects one of the possible denoting utterances with a probability proportional to the square of the interaction number stored with the general form used to produce the corresponding meaning. This choice means that repeatedly matched general forms will be selected with asymptotically uniform probability, while general forms that are only matched once or twice will be selected with probability tending to zero. This allows the learner gradually to abandon incorrect general forms as its understanding improves.

## 4.4 The learner: from meaning to utterance

A meaning is a sequence of variable atoms, for example,

$$(ci1(x_1), re1(x_1), le2(x_2, x_1), tr1(x_1)).$$

The learner processes this sequence to produce a sequence of words, that is, an utterance. A very basic approach is to process atoms in turn, replacing each atom with a word that approximately implies the atom's predicate. For this example, the result could be the utterance *circulo rojo derecha triangulo*, a kind of "telegraphic speech" that is understandable but not grammatically correct. A somewhat more complex strategy by the learner generally achieves grammatical correctness in our example domain. However, this strategy is clearly not sufficient in general for natural language.

A meaning is a sequence $(a_1, a_2, \ldots, a_k)$ of atoms. We associate with it two sequences of positions: the **atom positions** $1, 2, \ldots, k$ and the **gap positions** $0, 1, \ldots, k$. The atom positions refer to their corresponding atoms, and the gap position $i$ refers to the position to the right of atom $i$, except that gap position 0 is the position to the left of atom $a_1$. The learner generates an utterance by generating a sequence of zero or more words for each position of the meaning in left to right order, that is, first gap position 0, then atom position 1, then gap position 1, then atom position 2, and so on, until gap position $k$. The sequences of words generated are concatenated to form the final utterance.

In our system, the choice of what sequence of words to produce for each position in a meaning is represented by two sets of decision trees, for the atom positions and the gap positions.

For each variable atom that the learner has encountered in a unique teacher meaning, there is a decision tree that determines what sequence of words to produce for that atom in the context of the whole meaning. As an example, consider the atom $re1(x_1)$. In the sublanguage of Spanish represented by the transducer in Figure 1, this is unconditionally translated as *rojo*. However, in an extended language in which there are feminine nouns for shapes, the decision tree to decide how to translate $re1(x_1)$ branches on the value of the shape predicate applied to $x_1$ to select either *rojo* or *roja* as appropriate. These trees are used to determine the sequence of words produced for the atoms occurring in the atom positions of a word.

To handle the gap positions, for each generalization of a variable atom that has been encountered, there is a decision tree. The purpose of this decision tree is to determine the sequence of words produced for the gap position $i$ immediately following atom position $i$, where the atom in atom position $i$ matches the generalized atom associated with this decision tree. As in the case of

decision trees for atom positions, decision trees for gap positions branch on atoms in the meaning being translated to words. Gap position 0 does not follow any atom position, and has a separate decision tree.

For example, in an extended sublanguage of Spanish in which there are both masculine and feminine nouns for shapes, the learner learns a decision tree for a gap position following the atom $ab2(x_1, x_2)$ that branches on the value of the atom in the meaning that matches $shape1(x_2)$, choosing *del* for shapes denoted by masculine nouns and *de la* for shapes denoted by feminine nounds.

If there is no decision tree associated with a given atom or gap position in a meaning, the learner falls back on a "telegraphic speech" strategy as follows. For a gap position with no decision tree, no words are produced. For an atom position whose atom has no associated decision tree, the learner searches the transitive reduction of the approximate implication graph for words that approximately imply the predicate of the atom. If there are several such words, the one that has occurred in the most situations is chosen to replace the atom in the meaning.

## 4.5   The teacher's response to the learner's utterance

If the learner produces an utterance, the teacher analyzes it and then chooses its own utterance for the situation. The teacher may find the learner's utterance correct, incorrect but correctable, or incorrect and uncorrectable. In the case that the learner's utterance is found incorrect but correctable, the teacher chooses a possible correction for it. Then the teacher randomly chooses whether or not to use the correction as its utterance; a **correction probability** parameter governs this choice. If the learner produced no utterance or an uncorrectable one, of if the teacher did not choose to correct the learner's utterance, then the teacher's utterance is chosen uniformly at random from the denoting utterances for the situation. Thus, even when the learner's utterance is correct there is some probability that the teacher will simply repeat it.

The process used by the teacher to analyze the learner's utterance is as follows. If the learner's utterance is equal to one of the correct denoting utterances for the situation, the teacher classifies the learner's utterance as **correct**. If the learner's utterance is not correct, the teacher "translates" the learner's utterance into a sequence of predicates by using the adult meaning transducer for the language. In particular, each word for which there is a non-empty output in the transducer is replaced by the predicate from one such output. For example, the teacher translates the incorrect learner utterance *el elipse pequeno* into the predicate sequence $(el1, sm1)$ despite the errors of agreement in the utterance.

If the resulting sequence of predicates is the same as the sequence of predicates in at least one meaning obtained from a correct denoting utterance, the learner's utterance is classified as having an **error in form**. In this case, the learner is judged to have chosen a correct meaning but an incorrect form to express that meaning. The goal of the teacher is to choose a possible correction with the same meaning as that intended by the learner. A possible correction is chosen by the teacher by considering the set of denoting utterances whose meanings have the same sequence of predicates and choosing one such utterance using a measure of similarity to the learner's utterance. For example, if $(el1, sm1)$ corresponds to a denoting utterance for the situation, the teacher may choose *la elipse pequena* as a possible correction for *el elipse pequeno*.

If the learner's utterance is not correct and its corresponding sequence of predicates is not equal to the sequence of predicates for any denoting utterance for the situation, the teacher uses a measure of similarity between the sequence of predicates for the learner's utterance and the sequences of

predicates corresponding to denoting utterances for the situation to determine whether there is a "close enough" match between the predicate sequences. If so, the teacher classifies the learner's utterance as having an **error in meaning** and chooses as the possible correction a denoting utterance whose predicate sequence is judged "most similar" to the predicate sequence for the learner's utterance. For example, if the learner's utterance is *el pequeno* and the predicate sequence (*el1, sm1*) corresponds to a denoting utterance for the situation, the teacher may choose *la elipse pequena* as the possible correction.

If the learner produces an utterance but none of these cases (correct, error in form, error in meaning) apply, then the teacher classifies the learner's utterance as **uninterpretable** and does not offer a correction.

When the teacher has produced an utterance, the learner analyzes it and updates its grammar of the language as reflected in the co-occurrence graph, the general forms, and the decision trees for word choice. The decision trees are updated using a computed alignment between the teacher's utterance and the learner's understanding of the teacher's meaning, which assigns a (possibly empty) subsequence of words from the utterance to each gap or atom position in the meaning. Each subsequence of words is then added to the data for the corresponding decision tree.

If the learner has produced an utterance and finds that the teacher's utterance has the meaning intended by the learner, but is expressed differently, then the learner classifies the teacher's utterance as a **correction**. Of course, the learner may fail to recognize an intended correction of the teacher, and may also be mistaken when it classifies the teacher's utterance as a correction.

This analysis completes one interaction. A new situation is generated, and the cycle of learner production, teacher analysis, teacher production, and learner analysis is repeated with the new situation.

To illustrate the process, we provide a commented excerpt from the first fifty interactions of a learner and teacher in Appendix A. These interactions show the learner beginning to comprehend the teacher's utterances and acquiring and using both incorrect and correct general forms, producing both incorrect and correct denoting utterances. It also shows the teacher comprehending enough of the learner's meaning to offer meaning-preserving corrections, some of which the learner recognizes as corrections.

# 5    Empirical Results

We have implemented and tested our learning and teaching procedures in order to explore questions about the roles of semantics and corrections in language learning. The questions that we address in this section are the following. (1) Can the learner accomplish the learning task to a high level of correctness and coverage from a "reasonable" number of interactions (that is, well short of the number needed to memorize every legal situation/sentence pair)? (2) What are the effects of correction or non-correction by the teacher on the learner's accomplishment of the learning tasks? (3) Does the presence of semantics facilitate the learning process?

## 5.1    The learning tasks

The learning tasks we consider use the following set of situations. Each situation has two objects, each with three attributes (shape, color and size), and one binary relation between the two objects (above or to the left of.) The attribute of shape has six possible values (circle, square, triangle,

star, ellipse, and hexagon), that of color has six possible values (red, orange, yellow, green, blue, and purple), and that of size three possible values (big, medium, and small.) Thus there are 108 distinct objects and 23,328 distinct situations. Situations are generated uniformly at random.

We consider limited sublanguages of natural language utterances related to these situations for several natural languages. The utterances in a situation are phrases intended to denote one of the objects, for example, *the circle*, *the orange star*, or *the small purple hexagon below the medium green square*. In our languages, the shape attribute is expressed as a noun and must appear, while the size and color attributes are expressed as adjectives and may be omitted. If the binary relation between the objects is "above", it may be expressed using either above or below; if it is "to the left of", it may be expressed using either left or right. Thus, there are 168 meanings referring to a single object and 112,896 meanings referring to two objects, for a total of 113,064 possible meanings. The number of meanings that are denoting in a situation varies from 32 (for a situation with two objects with identical attributes) to 40 (for a situation with objects of two different shapes.)

In each of these languages, the 113,064 possible meanings are instances of 68 general forms: 4 referring to a single object and 64 referring to two objects. For English, examples of general forms are $(shape1(x_1))$, $(color1(x_1), shape1(x_1))$ and $(size1(x_1), shape1(x_1), le2(x_2, x_1), shape1(x_2))$. We refer to these languages as the **68-form languages**.

For each language we consider, we consulted at least one speaker of the language to help us construct a meaning transducer to translate appropriate phrases in the language to all 113,064 possible meanings. Each transducer was constructed to have exactly one accepted phrase for each possible meaning. Words were represented by ASCII strings, which necessitated using straight-forward transliterations of languages other than English. Examples of teacher utterances in the 68-form languages with some notes on their construction are in Appendix B. The procedure implementing the teacher is language-independent, and takes as parameters the meaning transducer for the language and the correction probability.

To help understand the effect of different aspects of the learning problem, we also considered reduced sublanguages, consisting of the utterances that refer to a single object (168 utterances) and those that refer to two objects, but include all three attributes of both (46,656 utterances.) Thus, for English the utterances in the reduced sublanguage include *the star*, *the blue triangle* and *the medium purple ellipse to the left of the medium red square* but not *the circle below the yellow hexagon*. Each meaning in the reduced sublanguage is an instance of one of 8 general forms, but most of the lexical and syntactic complexity of the language is preserved. We refer to these reduced sublanguages as the **8-form languages**.

## 5.2    How many interactions are needed to learn?

The level of performance of a learner attempting to learn a given language $L$ is measured using two quantities: the correctness and completeness of the learner's utterances in a given situation. The learning procedure is equipped with a test mode, in which no learning takes place. In the test mode, the learner receives a situation, and responds with all the utterances it could produce in that situation, and, for each one, the learner's probability of producing it. To evaluate the correctness and completeness of the learner's responses, the meaning transducer for $L$ is used to produce all the correct denoting utterances for the given situation. The **correctness** of the learner in the given situation is the sum of the probabilities of the learner's utterances that are in the correct denoting set. The **completeness** of the learner in the given situation is the fraction of the correct denoting utterances that appear in the set of learner utterances. The averages of

correctness and completeness of the learner in 200 randomly generated situations is used to estimate the overall correctness and completeness of the learner. We say that a learner reaches a **level $p$ of performance** if both correctness and completeness are at least $p$.

In our first set of trials, we set a target level of performance of $p = 0.99$ and the learner and teacher engage in a sequence of interactions until the learner first reaches this level of performance. Because the testing process is compute-intensive, the learner is tested only at intervals of 100 interactions. In Table 2 we show the number of interactions needed to first reach the given levels of performance for each 68-form language. Each entry is the median of 10 trials. For all these trials the teacher has a correction probability of 0.0.

In these results, there are two clear groups: one group is Greek and Russian, for which the median is at least 3400 interactions to reach level 0.99, and the other group is the rest of the languages, for which the median is at most 1000 interactions to reach level 0.99. The most basic observation is that the learner is able to achieve correctness and completeness of 0.99 for each of these languages after being exposed to a small fraction of all possible situations and utterances. In the case of Russian, 3700 interactions involve at most 16.5% of all possible situations and at most 3.5% of all possible utterances by the teacher. The bound of 1000 interactions represents fewer than 4.3% of all situations and fewer than 1% of all possible utterances.

To give an idea of how the learner progresses in a single trial, we show the learning curves for a typical run for Hebrew in Figure 3 and Russian in Figure 4. The horizontal axis is the number of interactions between learner and teacher. Two of the curves show the measurements of correctness and completeness of the learner's utterances every 100 interactions, tested on 200 random situations as described above. The third curve shows the fraction of the total number of general forms that have been acquired by the learner at intervals of 100 interactions. It is clear from these examples that neither correctness nor completeness is necessarily monotonically increasing.

To achieve completeness of at least 0.99, the learner must acquire all 68 correct general forms. Typically the learner also acquires a small number (3 or 4) of incorrect general forms in early interactions, whose probabilities decrease as they fail to be matched in later interactions. In both of these runs, the learner acquired its last general form between interaction 900 and interaction 1000. For the Hebrew run, the acquisition of the last general form essentially coincided with the achievement of the 0.99 level, while for the Russian run, levels of correctness and completeness continued to improve for another 2800 interactions after the last general form was acquired.

## 5.3  How do corrections affect learning?

We have seen (in the detailed analysis in Appendix A) that the teacher can detect and classify errors on the part of the learner, and offer corrections related to its interpretation of the learner's meaning. For syntactic errors (errors in form), the teacher's correction has the same meaning as its interpretation of the learner's meaning, and for semantic errors (errors in meaning), the teacher's correction has a meaning close to its interpretation of the learner's meaning. We have also seen that the learner can detect corrections intended by the teacher. The learner also occasionally classifies a teacher utterance as a correction when it is not intended as a correction. In this section we attempt to quantify some of the effects of corrections on the learning process.

**Corrections and 68-form languages.**  Table 3 shows the results of a set of trials parallel to those reported in Table 2, in which the teacher's correction probability is set to 1.0. This means

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|-----------|------|------|------|------|------|------|
| English | 200 | 200 | 300 | 400 | 500 | 700 |
| German | 200 | 300 | 300 | 400 | 550 | 800 |
| Greek | 400 | 500 | 700 | 1500 | 2200 | 3400 |
| Hebrew | 200 | 300 | 400 | 500 | 650 | 900 |
| Hungarian | 200 | 300 | 350 | 450 | 550 | 750 |
| Mandarin | 200 | 200 | 300 | 400 | 500 | 700 |
| Russian | 450 | 500 | 850 | 1750 | 2350 | 3700 |
| Spanish | 200 | 300 | 350 | 500 | 600 | 1000 |
| Swedish | 200 | 300 | 300 | 400 | 600 | 1000 |
| Turkish | 200 | 200 | 300 | 400 | 550 | 800 |

Table 2: Numbers of interactions needed to reach the given levels of performance with 68-form languages and probability 0.0 of correction. Each number is the median of 10 trials.

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|-----------|------|------|------|------|------|------|
| English | 200 | 200 | 300 | 400 | 500 | 750 |
| German | 200 | 300 | 400 | 500 | 500 | 750 |
| Greek | 400 | 500 | 700 | 1300 | 1700 | 2600 |
| Hebrew | 300 | 300 | 400 | 500 | 600 | 900 |
| Hungarian | 300 | 300 | 400 | 450 | 550 | 800 |
| Mandarin | 200 | 300 | 300 | 400 | 550 | 800 |
| Russian | 450 | 600 | 850 | 1500 | 2000 | 2900 |
| Spanish | 300 | 300 | 350 | 500 | 600 | 850 |
| Swedish | 200 | 300 | 300 | 500 | 600 | 900 |
| Turkish | 200 | 250 | 300 | 400 | 550 | 900 |

Table 3: Numbers of interactions needed to reach the given levels of performance with 68-form languages and probability 1.0 of correction. Each number is the median of 10 trials.
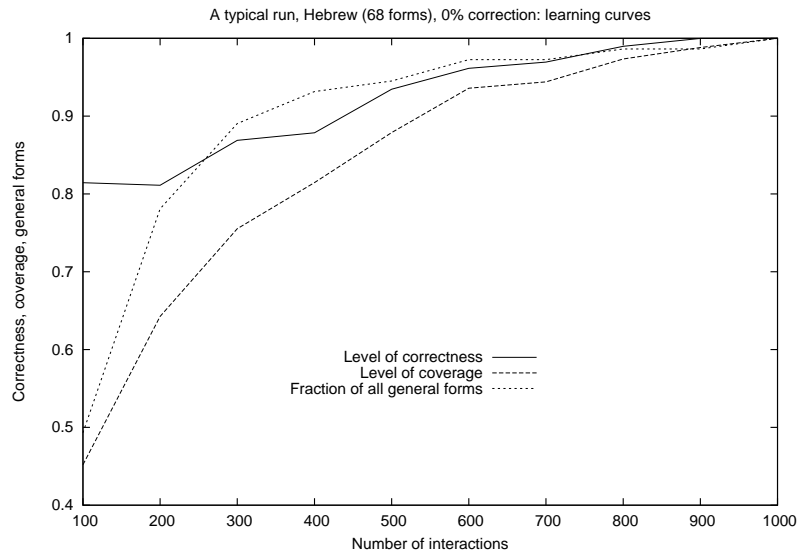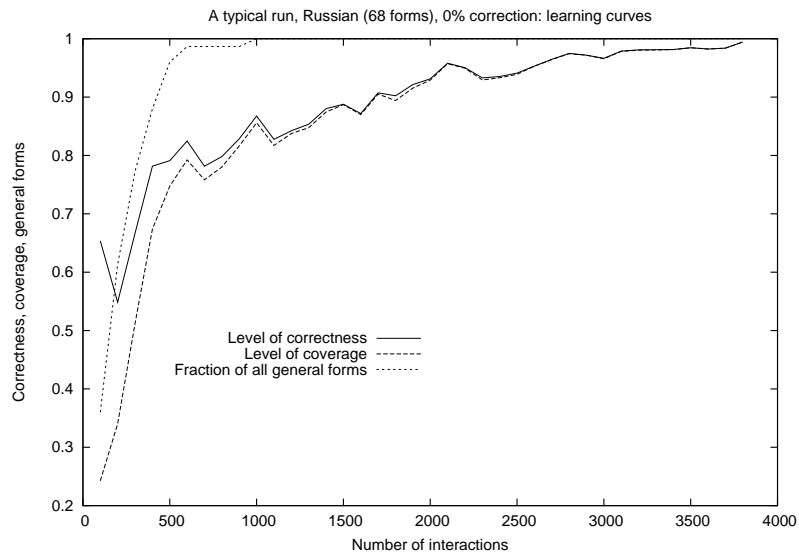
Figure 3: Learning curves for Hebrew



Figure 4: Learning curves for Russian

that the teacher offers a correction to the learner every time it classifies the learner's utterance as an error in form or an error in meaning.

These results fall into the same two groups: Greek and Russian versus the rest of the languages tested. For Greek, the median number of interactions to reach the 0.99 level of performance with correction probability 0.0 is 3400 and with correction probability 1.0 it is 2600, a decrease of about 24%. For Russian, the median with correction probability 0.0 is 3700 interactions and with correction probability 1.0 is 2900, a decrease of about 21%. Thus for these two languages there is a clear decrease in the number of interactions required to achieve level of performance 0.99. However, the languages in the other group show no clear effect of corrections in these trials.

It is important to note that in these trials the learner processes a teacher utterance in *exactly the same way* regardless of whether the learner classifies it as a correction. That is, the learner does not do anything special for perceived corrections. Thus, the improvements in performance for Greek and Russian depend entirely on the difference in behavior of the teacher. In particular, the non-correcting teacher and the correcting teacher produce different distributions on utterances, and this change in the learner's environment changes the performance of the learner. This suggests that studies of correction in natural languages should consider the possibility that corrections may affect the learning process *even when the learner cannot detect corrections (or does not attempt to.)*

To illuminate these results further, Table 4 presents the numbers of utterances, errors and corrections to reach level of performance 0.99 for the trials shown in Table 3. This shows that the teacher offers corrections for a substantial fraction of learner errors, and the fraction increases with increasing numbers of errors. The last column shows the percentage of the total number of teacher utterances that were corrections. Greek and Russian show the highest percentages of corrections, while English shows the lowest. The percentages for the other languages are intermediate; for these languages, we might expect that corrections could have a demonstrable (though more subtle) effect on learner performance.

|  | utterances | errors | corrections | percentage: c/u |
|---|---|---|---|---|
| English | 750 | 25.0 | 11.5 | 1.5% |
| German | 750 | 71.5 | 52.5 | 7.0% |
| Greek | 2600 | 344.0 | 319.0 | 12.3% |
| Hebrew | 900 | 89.5 | 62.5 | 6.9% |
| Hungarian | 800 | 76.5 | 58.5 | 7.3% |
| Mandarin | 800 | 50.0 | 31.5 | 3.9% |
| Russian | 2900 | 380.0 | 357.0 | 12.3% |
| Spanish | 850 | 86.0 | 68.0 | 8.0% |
| Swedish | 900 | 54.0 | 43.5 | 4.8% |
| Turkish | 900 | 59.0 | 37.0 | 4.1% |

Table 4: Cumulative numbers of interactions, errors and corrections at 0.99 level of performance for 68-form languages, with 1.0 probability of correction. Each number is the median of 10 trials. The last column gives the percentage that the corrections column is of the utterances column.

As the level of the learner's performance improves, there are several processes at work. The learner's comprehension of teacher utterances improves as more information about words and predicates accumulates in the co-occurrence graph. The learner acquires new correct general forms, and

earlier-acquired incorrect general forms decrease in probability as they are not matched by teacher utterances. The learner also acquires more accurate rules for choosing phrases for meanings that are instances of general forms. The attainment of the 0.99 level of performance may be limited by the need to acquire all the correct general forms (as seems to be the case for the Hebrew trial in Figure 3), or by the need to improve the correctness of the phrase choices (as seems to characterized the Russian trial in Figure 4.)

The difference in which process (form acquisition versus phrase choice improvement) is the bottleneck in attaining performance level 0.99 distinguishes the group consisting of Greek and Russian from the group consisting of the other languages in our sample. Information about the acquisition of general forms in these runs is shown in Table 5 for correction probability 0.0 and Table 6 for correction probability 1.0. In the case of Greek and Russian, most of the runs had acquired their last general form by the time the 0.90 level of performance was reached. The other languages show a more substantial change in the number of general forms between the 0.95 level and the 0.99 level in both conditions.

This data suggests that the bottleneck for Greek and Russian is the improvement of phrase choice, while for the other languages it is the acquisition of all 68 correct general forms. Because the teacher's corrections generally do not help with the acquisition of new general forms (in fact, very often the general form in a correction is the same one the learner just used), but do tend to improve the correctness of phrase choice, we do not expect correction to accelerate the attainment of the 0.99 level of performance when the bottleneck is the acquisition of general forms. This observation led us to construct reduced sublanguages with just 8 general forms to see if correction would have an effect when the bottleneck of acquiring general forms was removed. The next section describes results for these reduced sublanguages.

**Correction and 8-form languages.** The reduced sublanguages have just 8 general forms, which are acquired relatively early. In Table 7 we show the results of 100 trials each of the 8-form Spanish sublanguage, with correction probabilities of 0.0, 0.25, 0.50, 0.75 and 1.0. Levels of performance were tested every 50 interactions for these trials.

These results show an improvement of over 30% in the number of interactions to reach performance level 0.99 in going from correction probability 0.0 to correction probability 1.0. The intermediate correction probabilities give intermediate values. The corrected learners seem to be slightly slower to reach the 0.60 level of performance. The data on numbers of general forms for these runs show that learners had overwhelmingly acquired their last general form by the time they reached the 0.90 level of performance, where the superior performance of the corrected learners becomes very evident.

To give a more detailed sense of the comparison between the uncorrected and corrected learners, we include histograms of the numbers of interactions to reach the 0.99 level of performance in these trials for probability of correction 0.0 (in Figure 5) and 1.0 (in Figure 6.) The histograms are displayed on the same scale and convincingly show the advantage of the corrected learners in this task.

Table 8 gives the numbers of interactions to reach the given levels of performance for the 8-form sublanguages with correction probability 0.0. Table 9 gives the corresponding results for correction probability 1.0.

Comparing the results for 8-form sublanguages with corresponding 68-form languages, we see that some require notably fewer interactions to reach the 0.99 level of performance for 8-form sub-

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|
| English | 51.7 | 54.9 | 61.1 | 66.5 | 69.5 | 71.6 |
| German | 50.2 | 60.4 | 62.4 | 68.0 | 69.9 | 71.4 |
| Greek | 58.7 | 65.6 | 69.4 | 72.9 | 73.0 | 73.1 |
| Hebrew | 55.9 | 61.7 | 66.2 | 68.9 | 70.6 | 71.6 |
| Hungarian | 54.3 | 61.4 | 64.2 | 67.1 | 69.4 | 71.6 |
| Mandarin | 49.3 | 53.9 | 59.8 | 66.5 | 68.6 | 70.3 |
| Russian | 65.9 | 68.9 | 72.0 | 72.5 | 72.5 | 72.5 |
| Spanish | 53.9 | 59.6 | 63.4 | 67.9 | 70.1 | 71.1 |
| Swedish | 49.7 | 57.9 | 60.3 | 65.9 | 68.0 | 70.1 |
| Turkish | 51.1 | 55.3 | 60.4 | 65.8 | 69.3 | 71.1 |

Table 5: Cumulative number of general forms acquired at each level of performance for 68-form languages and probability 0.0 of correction. Each number is the mean of 10 trials.

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|
| English | 48.5 | 55.0 | 59.2 | 65.7 | 69.0 | 70.9 |
| German | 47.0 | 56.8 | 60.7 | 66.5 | 68.7 | 70.5 |
| Greek | 66.3 | 68.4 | 70.6 | 71.5 | 71.5 | 71.5 |
| Hebrew | 52.7 | 58.0 | 62.8 | 66.6 | 68.7 | 71.2 |
| Hungarian | 55.3 | 57.8 | 62.9 | 66.4 | 68.9 | 71.6 |
| Mandarin | 48.0 | 55.2 | 59.7 | 66.0 | 68.4 | 70.7 |
| Russian | 61.2 | 67.7 | 72.6 | 74.0 | 74.0 | 74.0 |
| Spanish | 51.3 | 56.1 | 59.2 | 64.3 | 67.0 | 69.3 |
| Swedish | 47.1 | 57.5 | 60.3 | 65.4 | 68.0 | 70.3 |
| Turkish | 47.1 | 53.7 | 58.3 | 65.2 | 67.9 | 70.3 |

Table 6: Cumulative number of general forms acquired at each level of performance for 68-form languages and probability 1.0 of correction. Each number is the mean of 10 trials.
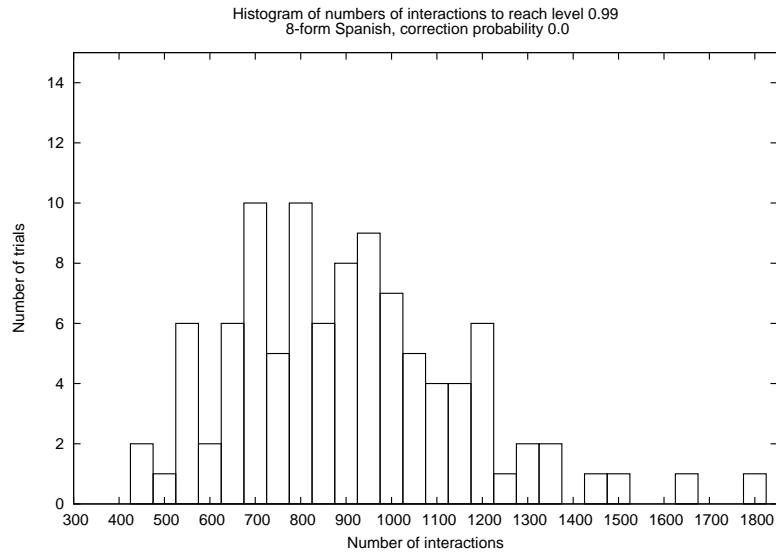
Figure 5: Histogram of number of interactions to achieve level 0.99 in 100 trials of learning 8-form Spanish with correction probability 0.0. Levels of performance were tested every 50 interactions.
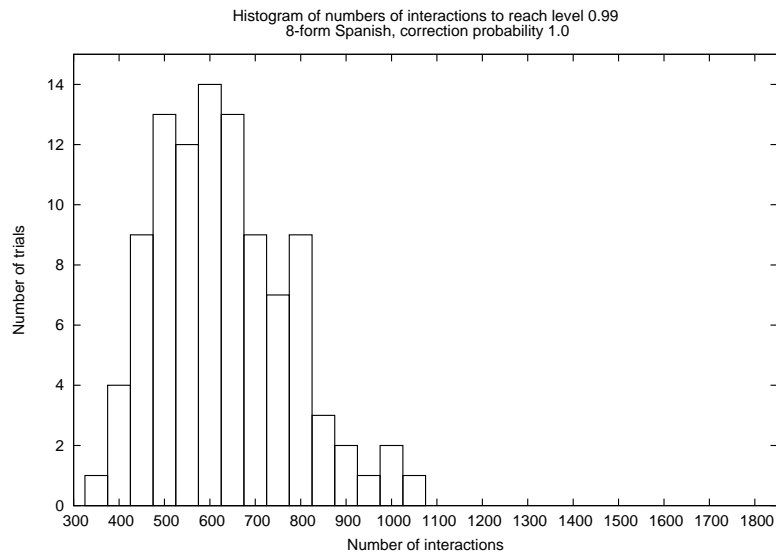


Figure 6: Histogram of number of interactions to achieve level 0.99 in 100 trials of learning 8-form Spanish with correction probability 1.0. Levels of performance were tested every 50 interactions.

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|-------|------|------|------|------|------|------|
| 0.00 | 86.5 | 128.0 | 200.5 | 359.5 | 541.0 | 908.0 |
| 0.25 | 91.0 | 123.0 | 186.0 | 316.5 | 472.5 | 841.5 |
| 0.50 | 90.0 | 125.5 | 189.0 | 297.5 | 438.5 | 730.5 |
| 0.75 | 92.5 | 121.0 | 169.5 | 283.0 | 421.0 | 674.0 |
| 1.00 | 93.5 | 122.0 | 170.5 | 280.0 | 386.5 | 630.5 |

Table 7: Numbers of interactions needed to reach the given levels of performance with 8-form Spanish and the indicated probabilities of correction. Levels of performance were tested every 50 interactions. Each number is the mean value for the 100 trials.

languages (English, Mandarin, and Swedish) while others require notably more (Greek, Hungarian and Russian.) In the case of Turkish, the learner cannot attain the 0.99 level of performance for the 8-form sublanguage at all, though it does so for the 68-form language, a result we explain further below. Thus, the 8-form languages are neither uniformly easier nor uniformly harder than their 68-form counterparts. Arguably, the restrictions that produce the 8-form languages make them "more artificial" than the 68-form languages; however, the artificiality helps us understand more about the possible roles of correction in language learning.

In Tables 8 and 9, there is no entry for performance level 0.99 for the case of the 8-form sublanguage of Turkish. Because of a limitation on the methods by which our learner comprehends the teacher's utterances and builds decision trees, it never "properly" understands teacher utterances that refer to two objects of the same shape, and therefore never adds them to the data for its decision trees. The distribution probabilities in the 8-form language mean that the default choice for the noun for the first of the two equal shapes in an utterance is in the wrong case. As a consequence, our learner cannot attain the 0.99 level of performance for the case of the 8-form sublanguage of Turkish though it can attain the 0.95 level of performance in this case. This is a useful reminder that our learning algorithm is in some important respects very crude. The inter-language comparisons that we give correlate generally with some properties of the underlying grammars, but also depend strongly on properties of our learning algorithm.

Even though in the case of the 8-form languages there are only 8 correct general forms to acquire, the distribution on utterances with one object versus utterances with two objects is quite different from the case of the 68-form languages. For a situation with two objects of different shapes, there are 40 denoting utterances in the case of 68-form languages, of which 8 refer to one object and 32 refer to two objects. In the case of the 8-form languagues, there are 10 denoting utterances, of which 8 refer to one object and 2 refer to two objects. Thus, in situations of this kind (which are 5/6 of the total), utterances referring to two objects are 4 times more likely in the case of 68-form languages than in the case of 8-form languages. This means that if the learner needs to see utterances involving two objects in order to master certain aspects of syntax (for example, cases of articles, adjectives and nouns), the waiting time is noticeably longer in the case of 8-form languages.

This longer waiting time emphasizes the effects of correction, because the initial phase of learning is a smaller fraction of the whole. In Table 10 we show the percentage reduction in the number of interactions to reach the 0.99 level of performance (except: 0.95 for Turkish) from correction probability 0.0 to correction probability 1.0 for the 8-form languages. For each language, corrections produce a reduction, ranging from a low of 11.7% for Swedish to a high of 38.1% for Greek.

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|
| English | 50.0 | 52.5 | 69.5 | 114.5 | 163.0 | 247.0 |
| German | 73.0 | 99.5 | 162.0 | 272.0 | 423.5 | 920.0 |
| Greek | 150.0 | 275.0 | 527.5 | 1532.5 | 3367.5 | 6630.0 |
| Hebrew | 70.0 | 120.0 | 203.5 | 342.0 | 504.5 | 1052.0 |
| Hungarian | 53.5 | 61.0 | 111.0 | 482.5 | 941.5 | 1632.5 |
| Mandarin | 61.0 | 78.0 | 99.5 | 163.0 | 202.5 | 340.5 |
| Russian | 82.5 | 215.0 | 542.5 | 1627.5 | 3932.5 | 6962.5 |
| Spanish | 86.5 | 128.0 | 200.5 | 359.5 | 541.0 | 908.0 |
| Swedish | 55.0 | 67.0 | 91.5 | 116.5 | 157.5 | 214.0 |
| Turkish | 58.0 | 69.0 | 161.0 | 466.5 | 1112.0 | **** |

Table 8: Numbers of interactions needed to reach the given levels of performance with 8-form languages and probability 0.0 of correction. Levels of performance were tested every 50 interactions. The numbers for Greek and Russian are the means of 20 trials; all others are the means of 100 trials. See the text for the explanation of the results for Turkish.

| Level | 0.60 | 0.70 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|
| English | 52.5 | 59.5 | 74.0 | 115.5 | 151.5 | 202.0 |
| German | 73.0 | 95.5 | 156.0 | 235.0 | 335.5 | 683.5 |
| Greek | 137.5 | 227.5 | 397.5 | 1017.5 | 2155.0 | 4102.5 |
| Hebrew | 74.0 | 118.0 | 183.0 | 280.5 | 405.0 | 771.5 |
| Hungarian | 57.5 | 65.0 | 110.5 | 348.0 | 637.5 | 1060.5 |
| Mandarin | 75.0 | 88.5 | 109.5 | 154.0 | 190.5 | 297.5 |
| Russian | 72.5 | 117.5 | 477.5 | 1130.0 | 2507.5 | 4640.0 |
| Spanish | 93.5 | 122.0 | 170.5 | 280.0 | 386.5 | 630.5 |
| Swedish | 54.0 | 67.0 | 85.5 | 109.5 | 144.0 | 189.0 |
| Turkish | 59.0 | 71.0 | 149.0 | 328.5 | 772.0 | **** |

Table 9: Numbers of interactions needed to reach the given levels of performance with 8-form languages and probability 1.0 of correction. Levels of performance were tested every 50 interactions. The numbers for Greek and Russian are the means of 20 trials; all others are the means of 100 trials. See the text for the explanation of the results for Turkish.

This confirms our hypothesis that corrections demonstrably help the learner when the problem of acquiring all the general forms is not the primary bottleneck.

| | 0.0 | 1.0 | % reduction |
|---|---|---|---|
| English | 247.0 | 202.0 | 18.2% |
| German | 920.0 | 683.5 | 25.7% |
| Greek | 6630.0 | 4102.5 | 38.1% |
| Hebrew | 1052.0 | 771.5 | 26.7% |
| Hungarian | 1632.5 | 1060.5 | 35.0% |
| Mandarin | 340.5 | 297.5 | 12.6% |
| Russian | 6962.5 | 4640.0 | 33.4% |
| Spanish | 908.0 | 630.5 | 30.6% |
| Swedish | 214.0 | 189.0 | 11.7% |
| Turkish | 1112.0 | 772.0 | 30.6% |

Table 10: Percentage reduction in the number of interactions to reach level 0.99 (except: 0.95 for Turkish) from probability of correction 0.0 to probability of correction 1.0 for 8-form sublanguages. Levels of performance were tested every 50 interactions. The numbers in the first two columns are means of 100 trials (except: 20 trials each for Greek and Russian.) The third column shows the percentage reduction from the first column to the second.

## 5.4   How much does semantics help?

In this section we investigate whether the presence of semantics facilitates the learning process. In order to do that, we compare the performance of our system to the ALERGIA algorithm.

**The ALERGIA algorithm.**   The ALERGIA algorithm is one of the most well known algorithms to identify stochastic deterministic regular languages from positive data. This problem has attracted a lot of attention from researchers in the field of Grammatical Inference. In the paradigm of identification in the limit, Gold proved that regular languages cannot be identified from positive data only, but they can be identified if both positive and negative data are provided [12]. Despite Gold's results, research in Grammatical Inference has been focused mainly on learning from positive data only, because in many applications negative data is not available to the learner. One solution to this problem is to learn probabilistic deterministic finite automata (PDFA). PDFA have been successfully applied in several domains including computational linguistics, machine learning, speech recognition, circuit testing, computational biology and machine translation [35,36]. Several algorithms for learning PDFA from positive data have been proposed [6,21,27,34]. The ALERGIA algorithm introduced by Carrasco and Oncina [6] is considered to be one of the most successful approaches to the problem of learning PDFA from a positive sample, and it has become the basis of comparisons for algorithms that have been subsequently proposed.

The ALERGIA algorithm is based on a state merging approach. From a given set of positive examples, a prefix tree automaton is constructed (which stores the sample set), and states with enough statistical evidence of equivalence are merged systematically according to a fixed order. The merging of states is controlled by a parameter $\alpha \in [0, 1]$. The algorithm stops when further

merging is not possible. A full description of the ALERGIA algorithm can be found in the paper of Carrasco and Oncina [6].

**A comparison of ALERGIA and our system.** In order to investigate the benefits of semantics to the learner, we use the ALERGIA algorithm to learn limited sublanguages of Spanish and English, specifically, the 68-form sublanguages of Spanish and English. We compare the number of examples needed to reach a high level of performance by the ALERGIA algorithm with the number of interactions needed by our system. We say that the ALERGIA algorithm achieves a high level of performance if the correctness and completeness of the learner are at least 0.99. To estimate the correctness of the learner, we calculate the percentage of incorrect utterances generated by the learner by asking the learner to produce 1000 utterances, and then checking the correctness of these utterances using the meaning transducer for that language. To estimate the completeness of the learner, we calculate the percentage of missing utterances by giving 1000 utterances to the learner (which are generated by using the meaning transducer for that language) and checking whether the learner can parse them or not.

The utterances given to the ALERGIA algorithm are generated in the same way as the utterances of a non-correcting teacher. That is, to get a sequence of $n$ utterances to use as input to the ALERGIA algorithm, we generate a sequence of $n$ pairs of a random situation and a randomly chosen denoting utterance, and extract the sequence of utterances to form the input to ALERGIA. For each length $n = 1000, 2000, 3000, \ldots$, we independently generate 10 such sequences, and consider the median values of these 10 runs.

The parameter $\alpha$ controls the process of generalization in ALERGIA. Depending on how we fix this parameter, the number of sentences needed by the algorithm may vary substantially; if $\alpha$ is close to 0, ALERGIA generalizes aggressively and generates very small automata (one state in the limiting case), and if $\alpha$ is close to 1, ALERGIA simply memorizes the data. Thus, we must find an intermediate value for ALERGIA to work properly. So far, the only known way to choose a value fo $\alpha$ is experimentally. In order to do a fair comparison of our system with ALERGIA, we have conducted a number of experiments to try to find a good value of $\alpha$. The value has been adjusted to try to minimize the percentages of incorrect and missing sentences of the learner (in order that the percentages of incorrectness and missing are less than 1%).) Examples of the variation of the percentages of incorrect and missing utterances for different values of $\alpha$ are shown in Table 11. In these results, ALERGIA takes more examples to produce correct utterances than to correctly parse all the sentences given to it; by contrast in our system the learner generally achieves correctness before completeness.

After an extensive search, the best values we could find for the Spanish 68-form language was 0.05 and for the English 68-form language was 0.20. Using these values for $\alpha$, we ran the ALERGIA algorithm to estimate the minimum number of utterances required to acquire these two sublanguages.

In the case of Spanish, instead of 1000 interactions, ALERGIA needs 6000 utterances. For English, the difference is still larger; instead of 700 interactions, ALERGIA needs 15000 utterances. These results show that the number of examples is drastically reduced when semantics is taken into account, suggesting that the presence of semantics significantly facilitates the learning process. These results are graphically displayed in Figure 7.

In these results, ALERGIA needs many more examples to learn the 68-form English language than the 68-form Spanish language. In the case of our system, the opposite occurred, although

| $\alpha$ | % Incorrect | % Missing |
|---------|------------|-----------|
| 0.00005 | 9.35 | 0 |
| 0.0005 | 8.20 | 0 |
| 0.005 | 6.85 | 0 |
| 0.05 | 0.70 | 0 |
| 0.1 | 1.15 | 0.2 |
| 0.2 | 3.10 | 0.4 |
| 0.3 | 2.85 | 0.3 |
| 0.4 | 5.00 | 0.35 |
| 0.5 | 5.35 | 0.65 |
| 0.6 | 9.25 | 1.15 |
| 0.7 | 12.55 | 1.55 |
| 0.8 | 17.30 | 3.35 |
| 0.9 | 21.20 | 5.7 |

Table 11: Percentages of incorrect and missing utterances of ALERGIA learner with different values of the parameter $\alpha$ after learning from 6000 utterances in the 68-form Spanish language. Each number is the median of 10 trials.
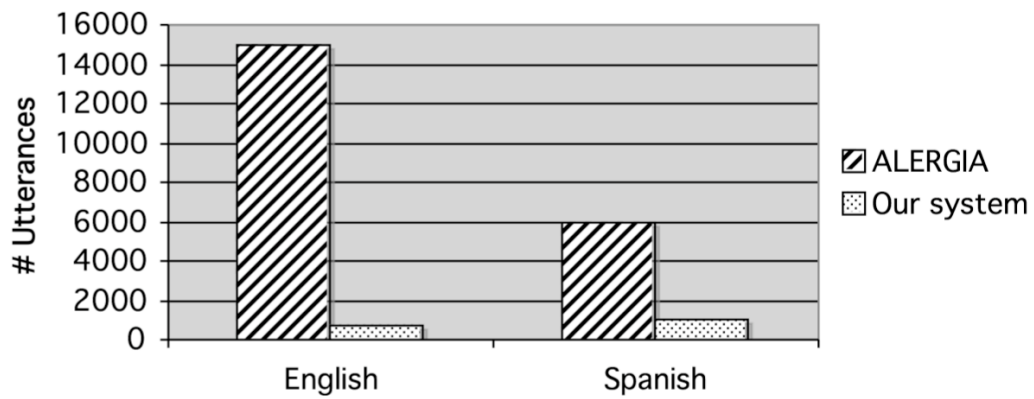


Figure 7: Sample size needed to reach the 0.99 level of performance learning Spanish and English 68-form languages, by the ALERGIA algorithm and our system.

the difference in the number of interactions was much smaller (just 300 interactions). One possible explanation of this phenomenon is that the branching factor near the root of the initial prefix tree of data is much smaller in the case of the Spanish 68-form language (*el* followed by four possible nouns and *la* followed by two possible nouns) than in the case of the English 68-form language (*the* followed by nine possible adjectives and six possible nouns), leading to a much more concentrated accumulation of statistical information in the Spanish case.

We note that for this comparison, we use the results of our system with probability 0.0 of correction. Our results in Section 5.3 suggest that if in addition to semantics, corrections are taken into account (e.g., the teacher's correction probability is set to 1.0), the improvement for some learning tasks could be even greater.

# 6    Collecting Corrections

In this section we define and analyze a simplified model of the interactions between learner and teacher to get a better understanding of the possibilities and limitations of corrections in our setting. We model the process of learning with corrections as a modified version of the coupon collector process.

In the basic coupon collector process, there is a set of $n$ tokens, each bearing a distinct positive integer from 1 to $n$. An agent draws tokens uniformly at random from this set with replacement, until it has drawn each number from 1 to $n$ at least once. The expected number of draws until the agent has drawn each distinct token at least once is $nH(n)$, where $H(n)$ is the $n^{th}$ harmonic number, that is

$$H(n) = \sum_{k=1}^{n} 1/k.$$

Asymptotically this quantity is $n \ln n + O(n)$.

To model learning with corrections in our setting, we define a **correction collection** process, as follows. The learner has a set of $n$ tokens, each bearing a distinct positive integer from 1 to $n$ and also a **correctness bit** indicating whether the token is correct or not. The teacher has a similar set of tokens, except that each correctness bit is set to 1. For each token, the learner can only perceive the number on the token, not its correctness bit, while the teacher can perceive both the number on the token and its correctness bit. The tokens represent possible utterances of the learner and the teacher. We assume that the learner cannot determine the correctness of utterances for itself but that the teacher can. Moreover we assume that the teacher's utterances are all correct at the start.

Now the learner and teacher engage in a sequence of interactions as follows. The learner draws one of its tokens uniformly at random with replacement and shows it to the teacher. The teacher chooses with replacement one of its tokens and shows it to the learner. If the teacher's token is numbered $i$, then the correctness bit on the learner's token numbered $i$ is set to 1, indicating that it is now correct. This is a simplified model of the interactions in our setting: the learner produces an utterance, then the teacher produces an utterance, and the learner uses the teacher's utterance as a model for correct utterances of that kind in the future.

We consider two possible strategies for the teacher's choice of a token to show to the learner. The non-correcting teacher ignores the token chosen by the learner and chooses one of its tokens uniformly at random to show to the learner. The correcting teacher looks at the token chosen by

29

the learner and chooses its action based on the correctness bit. If the learner's token has correctness bit 0 and number $i$, then the teacher chooses its own token with number $i$, thereby ensuring that the learner's token $i$ will henceforth be correct. If the learner's token has correctness bit 1, then the teacher chooses one of its own tokens uniformly at random. These two strategies model the actions of the teacher in our setting with correction probability 0.0 (when a random denoting utterance is chosen) or correction probability 1.0 (when the teacher chooses a correcting utterance whenever it can.)

We assume that all of the learner's tokens initially have correctness bit 0 and analyze the expected number of interactions until all of the learner's tokens have correctness bit 1. In the case of the non-correcting teacher, the process reduces immediately to the coupon collection process: the choices of the learner are irrelevant, and each learner's token has its correctness bit changed from 0 to 1 the first time the teacher randomly draws its own token with the same number. Thus in this case, the expected number of interactions until the learner's tokens are all corrected is $n \ln n + O(n)$.

In the case of the correcting teacher, we define the random variable $Y_c$ to be the number of the interaction in which the number of correct learner tokens first reaches $c$, with $Y_0 = 0$ and $Y_1 = 1$. We are interested in $E(Y_n)$. We define $X_c = Y_{c+1} - Y_c$ for $c = 0, 1, \ldots, n-1$. If the learner currently has $c$ correct tokens, then in one interaction it can come to have $c + 1$ correct tokens in two ways: by choosing one of its $(n - c)$ incorrect tokens and being corrected by the teacher, or by choosing one of its $c$ correct tokens and then having the teacher randomly choose a token corresponding to one of the learner's $(n - c)$ incorrect tokens, which is then corrected. Thus, the probability $p_c$ of moving from $c$ correct tokens to $c + 1$ tokens in one interaction is

$$p_c = \frac{(n - c)}{n} + \frac{c}{n} \frac{(n - c)}{n} = \frac{n^2 - c^2}{n^2}.$$

Then we may bound $E(Y_n)$ as follows.

$$E(Y_n) = \sum_{c=0}^{n-1} E(X_c)$$
$$= \sum_{c=0}^{n-1} \frac{1}{p_c}$$
$$= n^2 \sum_{c=0}^{n-1} \frac{1}{n^2 - c^2}$$
$$\leq n^2 \left( \int_0^{n-1} \frac{dx}{n^2 - x^2} \right) + \frac{n^2}{2n - 1}$$
$$< \frac{1}{2} n \ln n + O(n).$$

Asymptotically, the expected number of interactions until no uncorrected tokens remain for the learner interacting with the correcting teacher is $1/2$ of the value for the learner interacting with the non-correcting teacher. Despite the simplicity of this model, these results are roughly consistent with the rates of improvement seen in our empirical data for learners interacting with teachers correcting with probability 1.0, none of which exceeded about 40%.

This simplified model could be generalized to allow intermediate probabilities of correction by the teacher, and perhaps more complicated effects such as differential attention paid by the learner

to perceived corrections. This model highlights the fact that neither teacher nor learner takes into account the history of their interactions, relying on the fact that random choices cover the whole domain with reasonable efficiency.

# 7 Discussion and Future Work

Our model and results show that semantics can significantly reduce the number of examples needed to acquire limited sublanguages of natural language. This improvement depends on the learner's prior knowledge of the semantic domain and categories. The learner collects information about the co-occurrence of words and predicates, and uses the semantic situation to help it translate the teacher's utterances to meanings. The teacher's meanings are generalized via the semantic categories to general forms that are the basis of the learner's own utterances. The learner refines its ability to use the general forms to produce grammatical utterances by collecting data from alignments of the teacher's utterances with the general forms. This data is used to construct decision trees to choose sequences of words for parts of an intended meaning. Thus, the learner's grammar is represented by its initial semantic categories and the information it has collected about co-occurrence of words and predicates, about what general forms represent meanings, and about how to choose sequences of words for parts of an intended meaning.

Decomposing the learner's task into learning a grammar of meanings (in our system, the general forms) and a set of rules for expressing meanings in words (in our system, the decision trees) seems to simplify the learning task. However, it is likely that more complex language learning tasks require a more complex model of the relationship between meanings and utterances. For the tasks we consider, simply accumulating a list of general forms (with a mechanism to "age out" incorrect ones) and translating the atoms and gaps of the meaning into consecutive sequences of words provides a feasible solution.

Our model of language is very simplified, and there are many issues it does not deal with properly, including multi-word phrases bearing meaning, morphological relations between words, phonological rules for word choice, words with more than one meaning and meanings that can be expressed in more than one way, languages with free word-orders and meaning components expressed by non-contiguous sequences of words. Questions of generality, scalability and noise tolerance should also be addressed. Other desirable directions to explore include more sophisticated use of co-occurrence information, more powerful methods of learning the grammars of meanings, feedback to allow the learning of production to improve comprehension, better methods of alignment between utterances and meanings, and methods to allow the learner's semantic categories to evolve in response to language learning.

Our model and results have also allowed us to demonstrate that a relatively simple model of a teacher can offer meaning-preserving corrections to the learner, and that for certain learning tasks such corrections can significantly reduce the number of interactions for the learner to reach a high level of performance. Moreover, this improvement does not depend on the learner's ability to detect corrections: the effect depends on the change in the distribution of teacher utterances in the correcting versus non-correcting conditions. This result suggests re-visiting discussions in linguistics that assume that the learner must detect a teacher correction as a correction in order for it to have an influence on the learning process.

In fact, in our model the learner can detect meaning-preserving corrections of the teacher. A future direction of research is to see if a learner can use the ability to detect corrections to

accelerate the learning process further. Preliminary experiments in assigning greater weight to detected corrections in the decision tree construction process did not seem to have a significant positive effect on the attainment of high levels of performance.

Our model of a teacher is very simple and purely "reactive." That is, the teacher does not keep any history for the learner that it is interacting with (other than accumulating summary statistics), and reacts to learner utterances using the same procedures each time. The only dimension along which teachers differ is their propensity to correct the learner (that is, the probability of correction.) An interesting future direction would be to explore the effects of making the teacher "more responsive" to the learner, possibly taking into account the history of interactions with this learner. One possibility would be to try a simple model of "motherese" in which the distribution on the teacher's utterances might be weighted toward utterances slightly more complex than the current average for learner utterances. Another possibility would be to have the teacher keep track of learner errors, and weight its distribution of utterances to provide more models of correct utterances in regions of learner error.

It would be very interesting to extend our results to more practical domains such as data mining and information retrieval. In the framework of data mining, semantic information might allow us to tackle the problem of scalability by reducing the amount of data the algorithms must consider. In information retrieval, specifically in the context of web search, a search guided by semantic information might allow users to get more relevant results to their queries when the desired information is hard to find purely syntactically. Semantics might also facilitate communication between the user and the search engine; the user could ask for information in a more natural way (not just using keywords), and the search engine could understand the query, even if it is not completely grammatically correct. We believe that incorporating semantics is crucial for future search engines.

# 8  Acknowledgments

# References

[1] D. Angluin. Inference of reversible languages. *Journal of the Association for Computing Machinery*, 29(3):741–765, 1982.

[2] D. Angluin and L. Becerra-Bonache. Learning meaning before syntax. Technical Report YALE/DCS/TR1407, Computer Science Department, Yale University, 2008.

[3] D. Angluin and L. Becerra-Bonache. Learning meaning before syntax. In *ICGI 2008: 9th International Colloquium on Grammatical Inference*, pages 1–14. Springer, Berlin, 2008.

[4] D.R. Bailey. *A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. PhD thesis, U.C. Berkeley, 1997.

[5] R. Brown and C. Hanlon. Derivational complexity and the order of acquisition in child speech. In J.R. Hayes, editor, *Cognition and the Development of Language*. Wiley, New York, 1970.

[6] R.C. Carrasco and J. Oncina. Learning stochastic regular grammars by means of a state merging method. In *Proc. of ICGI 94: 2nd Int. Colloquium on Grammatical Inference*, pages 139–152, 1994.

[7] M.M. Chouinard and E.V. Clark. Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30:637–669, 2003.

[8] E.V. Clark. The principle of contrast: a constraint on language acquisition. In B. MacWhinney, editor, *Mechanisms of language acquisition*. Erlbaum, Hillsdale, NJ, 1987.

[9] E.V. Clark. *The Lexicon in Acquisition*. Cambridge University Press, Cambridge, 1993.

[10] C.G. de Marcken. *Unsupervised Language Acquisition*. PhD thesis, MIT, 1996.

[11] M. J. Demetras, K. N. Post, and C. E. Snow. Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language*, 13:275–292, 1986.

[12] E.M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[13] K. Gold. *Using Sentence Context and Implicit Contrast to Learn Sensor-Grounded Meanings for Relational and Deictic Words: The TWIG System*. PhD thesis, Yale University, 2008.

[14] K. Gold, M. Doniec, C. Crick, and B. Scassellati. Robotic vocabulary building using extension inference and implicit contrast. *Artifical Intelligence*, 173 (1):145–166, 2004.

[15] K. Gold and B. Scassellati. A robot that uses existing vocabulary to infer non-visual word meanings from observation. In *AAAI*, pages 883–888, 2007.

[16] J. A. Collins Hill. *A Computational Model of Language Acquisition in the Two-Year-Old*. Indiana University Linguistics Club, Indiana, 1983.

[17] K. Hirsh-Pasek, R. A. Treiman, and M. Schneiderman. Brown and Hanlon revisited: mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, 2:81–88, 1984.

[18] A. Stolcke J.A. Feldman, G. Lakoff and S. Weber. Miniature language acquisition: A touchstone for cognitive science. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, pages 686–693. MIT, Cambridge, MA, 1994.

[19] K. Jack. *A Computational Model of Staged Language Acquisition*. PhD thesis, University of Dundee, 2006.

[20] M. Kanazawa. *Learnable Classes of Categorial Grammars*. Cambridge University Press, New York, NY, 1998.

[21] C. Kermorvant and P. Dupont. Stochastic grammatical inference with multinomial tests. In *Proc. of ICGI 02: 6th Int. Colloquium on Grammatical Inference*, pages 149–160, 2002.

[22] G.F. Marcus. Negative evidence in language acquisition. *Cognition*, 46:53–95, 1993.

[23] J.L. Morgan and L.L. Travis. Limits on negative information in language input. *Journal of Child Language*, 16:531–552, 1989.

[24] Martin Redington, Nick Chater, and Steven Finch. Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22:425–469, 1998.

[25] T. Regier. *The Human Semantic Potential: Spatial Language and Constrained Connectionism*. MIT Press, Cambridge, MA, 1996.

[26] T. Regier. The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29:819–865, 2005.

[27] D. Ron, Y. Singer, and N. Tishby. On the learnability and usage of acyclic probabilistic finite automata. In *COLT 95: 8th Conference on Computational Learning Theory*, pages 31–40, 1995.

[28] D.K. Roy and A.P. Pentland. Learning words from sights and sounds: A computational model. *Cognitive Science*, 26:113–146, 2002.

[29] Y. Sakakibara. Efficient learning of context-free grammars from positive structural examples. *Information Processing Letters*, 97:23–60, 1992.

[30] A.M. Shaerlaekens. *The Two-Word Sentence in Child Language Development: A Study Based on Evidence Provided by Dutch-Speaking Triplets*. Mouton, The Hague, 1973.

[31] J.M. Siskind. Lexical acquisition in the presence of noise and homonymy. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI 94)*. MIT Press, 1994.

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

[33] I. Tellier. Meaning helps learning syntax. In *Proc. of ICGI 98*, pages 25–36, 1998.

[34] F. Thollard, P. Dupont, and C. de la Higuera. Probabilistic DFA inference using Kullback-Leibler divergence and minimality. In *ICML 2000: 17th International Conference on Machine Learning*, pages 975–982, 2000.

[35] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, and R. C. Carrasco. Probabilistic finite-state machines – part II. *IEEE Trans. Pattern Analysis Machine Intelligence*, 27(7):1026–1039, 2005.

[36] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, and R.C. Carrasco. Probabilistic finite-state machines – part I. *IEEE Trans. Pattern Analysis Machine Intelligence*, 27(7):1013–1025, 2005.

[37] C. Yu and D.H. Ballard. A multimodal learning interface for grounding spoken language in sensory perceptions. *ACM Transactions on Applied Perceptions*, 1 (1):57–80, 2004.

# A    Appendix: excerpted interactions

For illustration we include a commented selection from the first fifty interactions in one run of a learner learning 68-form Spanish from a teacher using correction probability 1.0.

```
> (test-x-sp 1 50 1.0)
(
((interaction-number 0)
(situation ((sm1 t1) (pu1 t1) (el1 t1) (le2 t1 t2) (bi1 t2) (pu1 t2) (tr1 t2)))
(learner-intended-meaning ())
(learner-utterance ())
(teacher-classification no-utterance)
(intended-correction? #f)
(teacher-utterance (la elipse purpura a la izquierda del triangulo))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)
((interaction-number 1)
(situation ((me1 t1) (pu1 t1) (ci1 t1) (le2 t1 t2) (sm1 t2) (or1 t2) (he1 t2)))
(learner-intended-meaning ())
(learner-utterance ())
(teacher-classification no-utterance)
(intended-correction? #f)
(teacher-utterance (el circulo purpura a la izquierda del hexagono pequeno y naranja))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...  For the first few interactions the learner
     produces no utterance while collecting
     information about the language.


((interaction-number 4)
(situation ((sm1 t1) (gr1 t1) (tr1 t1) (ab2 t1 t2) (sm1 t2) (bl1 t2) (he1 t2)))
(learner-intended-meaning ())
(learner-utterance ())
(teacher-classification no-utterance)
(intended-correction? #f)
(teacher-utterance (el triangulo verde encima del hexagono azul))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (shape1 x2)))
(learner-understanding ((tr1 x1) (he1 x2)))
)

... In this interaction, the learner acquires its first
    general form: ((shape1 x1) (shape1 x2)), which is
    incorrect.  However, it now attempts to use the
    form to produce denoting utterances.
```

```
((interaction-number 6)
(situation ((bi1 t1) (gr1 t1) (ci1 t1) (ab2 t1 t2) (sm1 t2) (gr1 t2) (el1 t2)))
(learner-intended-meaning ((el1 x1) (ci1 x2)))
(learner-utterance (el elipse verde encima del mediana azul))
(teacher-classification uninterpretable)
(intended-correction? #f)
(teacher-utterance (la elipse verde debajo del circulo grande))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...   The form is incorrect and the word choice rules are
      wrong, but the learner produces its first utterance,
      with "elipse" reflecting part of its intended meaning.
      The teacher finds the utterance uninterpretable.

((interaction-number 7)
(situation ((me1 t1) (re1 t1) (tr1 t1) (le2 t1 t2) (sm1 t2) (re1 t2) (st1 t2)))
(learner-intended-meaning ((st1 x1) (tr1 x2)))
(learner-utterance (el estrella verde encima del triangulo azul))
(teacher-classification uninterpretable)
(intended-correction? #f)
(teacher-utterance (la estrella pequena y roja a la derecha del triangulo mediano))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...   The learner continues using the incorrect form --
      "estrella" and "triangulo" correspond to its intended
      meaning.  The result is uninterpretable, and the teacher
      produces a new utterance.  Several similar exchanges ensue.

((interaction-number 14)
(situation ((me1 t1) (pu1 t1) (sq1 t1) (le2 t1 t2) (bi1 t2) (re1 t2) (el1 t2)))
(learner-intended-meaning ((el1 x1) (sq1 x2)))
(learner-utterance (el elipse verde encima del cuadrado azul))
(teacher-classification uninterpretable)
(intended-correction? #f)
(teacher-utterance (el cuadrado a la izquierda de la elipse grande))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (le2 x1 x2) (color1 x2) (shape1 x2) (size x2)))
(learner-understanding ((sq1 x1) (le2 x1 x2) (re1 x2) (el1 x2) (bi1 x2)))
)

...   The learner acquires another incorrect general form,
      apparently because it is confused about the meaning
      of "de la."  (The color predicate is misplaced.)

((interaction-number 15)
```

36

```
(situation ((bi1 t1) (or1 t1) (el1 t1) (ab2 t1 t2) (bi1 t2) (ye1 t2) (he1 t2)))
(learner-intended-meaning ((el1 x1) (he1 x2)))
(learner-utterance (el elipse hexagono))
(teacher-classification error-in-meaning)
(intended-correction? #t)
(teacher-utterance (la elipse encima del hexagono))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)
```

...   The learner continues using the first incorrect form,
      but with a greater understanding of the meanings of
      words.  The teacher can now understand the utterance
      enough to decide it is an error in meaning and offer
      a correction close in meaning to the learner's intention.
      The learner, however, does not detect it as a correction.

```
((interaction-number 16)
(situation ((bi1 t1) (pu1 t1) (ci1 t1) (le2 t1 t2) (me1 t2) (re1 t2) (ci1 t2)))
(learner-intended-meaning ((ci1 x1) (le2 x1 x2) (re1 x2) (ci1 x2) (me1 x2)))
(learner-utterance (el circulo a la izquierda de la circulo mediana))
(teacher-classification error-in-form)
(intended-correction? #t)
(teacher-utterance (el circulo a la izquierda del circulo mediano))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)
```

...   The learner again uses the second incorrect form, but
      because of confusion about the meanings of words
      (apparently the re1 predicate is rendered as "de la",)
      the result accidentally has a correct meaning, and
      the teacher classifies it as an error in form and
      gives a correction.  The learner does not detect a
      correction.

```
((interaction-number 18)
(situation ((sm1 t1) (bl1 t1) (tr1 t1) (le2 t1 t2) (bi1 t2) (bl1 t2) (st1 t2)))
(learner-intended-meaning ((st1 x1) (tr1 x2)))
(learner-utterance (el estrella triangulo))
(teacher-classification error-in-meaning)
(intended-correction? #t)
(teacher-utterance (la estrella a la derecha del triangulo))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (le2 x2 x1) (shape1 x2)))
(learner-understanding ((st1 x1) (le2 x2 x1) (tr1 x2)))
)
```

...   The learner uses the incorrect first general form,

receives a correction from the teacher, and acquires
        a third, correct, general form from the correction,
        which, however, the learner does not detect as a
        correction.   The learner continues using all
        three general forms, although the incorrect ones
        will gradually lose probability because they do
        not continue to be matched by teacher utterances.

((interaction-number 21)
(situation ((sm1 t1) (bl1 t1) (tr1 t1) (ab2 t1 t2) (me1 t2) (bl1 t2) (tr1 t2)))
(learner-intended-meaning ())
(learner-utterance ())
(teacher-classification no-utterance)
(intended-correction? #f)
(teacher-utterance (el triangulo mediano debajo del triangulo azul))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...   Even with three possible general forms, the learner
      may not be able to find a denoting meaning in some
      situations.  Here it can produce no utterance.

((interaction-number 22)
(situation ((me1 t1) (re1 t1) (st1 t1) (le2 t1 t2) (bi1 t2) (or1 t2) (st1 t2)))
(learner-intended-meaning ((st1 x1) (le2 x2 x1) (st1 x2)))
(learner-utterance (el estrella a la derecha del estrella))
(teacher-classification error-in-form)
(intended-correction? #t)
(teacher-utterance (la estrella a la derecha de la estrella))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

... Here the learner uses the correct third form for the
    first time, and the teacher classifies it as an
    error in form and offers a correction.  Because
    of the difficulty with left/right and above/below
    in symmetrical situations, the learner fails to
    recognize the correction.

((interaction-number 23)
(situation ((sm1 t1) (gr1 t1) (tr1 t1) (ab2 t1 t2) (sm1 t2) (pu1 t2) (el1 t2)))
(learner-intended-meaning ((el1 x1) (tr1 x2)))
(learner-utterance (el elipse triangulo))
(teacher-classification error-in-meaning)
(intended-correction? #t)
(teacher-utterance (la elipse debajo del triangulo))
(learner-detected-correction? #f)

```
(new-general-form ((shape1 x1) (ab2 x2 x1) (shape1 x2)))
(learner-understanding ((el1 x1) (ab2 x2 x1) (tr1 x2)))
)


...   The learner acquires a correct fourth general form
      from the teacher's undetected correction of the error
      in meaning caused by the incorrect first form.  An
      incorrect form again facilitates learning a correct form.


((interaction-number 24)
(situation ((me1 t1) (gr1 t1) (tr1 t1) (le2 t1 t2) (me1 t2) (gr1 t2) (tr1 t2)))
(learner-intended-meaning ((tr1 x1) (le2 x2 x1) (tr1 x2)))
(learner-utterance (la triangulo a la derecha del triangulo))
(teacher-classification error-in-form)
(intended-correction? #t)
(teacher-utterance (el triangulo a la derecha del triangulo))
(learner-detected-correction? #t)
(new-general-form ())
(learner-understanding ((tr1 x1) (le2 x2 x1) (tr1 x2)))
)


...   The learner for the first time detects a correction
      intended by the teacher.  The learner does not use
      the fact that it detected a correction; the teacher's
      utterance is processed the same way regardless.


((interaction-number 25)
(situation ((bi1 t1) (gr1 t1) (tr1 t1) (ab2 t1 t2) (bi1 t2) (pu1 t2) (el1 t2)))
(learner-intended-meaning ((el1 x1) (ab2 x2 x1) (tr1 x2)))
(learner-utterance (la elipse debajo del triangulo))
(teacher-classification correct)
(intended-correction? #f)
(teacher-utterance (el triangulo grande encima de la elipse purpura))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)


...   The learner for the first time produces an utterance
      classified as correct by the teacher.


((interaction-number 26)
(situation ((sm1 t1) (pu1 t1) (he1 t1) (le2 t1 t2) (sm1 t2) (re1 t2) (el1 t2)))
(learner-intended-meaning ((he1 x1) (le2 x1 x2) (re1 x2) (el1 x2) (sm1 x2)))
(learner-utterance (el hexagono a la izquierda de la elipse amarillo))
(teacher-classification error-in-meaning)
(intended-correction? #t)
(teacher-utterance (el hexagono a la izquierda de la elipse roja))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (le2 x1 x2) (shape1 x2) (color1 x2)))
```

```
(learner-understanding ((he1 x1) (le2 x1 x2) (el1 x2) (re1 x2)))
)

...  The learner acquires a correct fifth general form.

((interaction-number 30)
(situation ((bi1 t1) (pu1 t1) (he1 t1) (le2 t1 t2) (sm1 t2) (or1 t2) (sq1 t2)))
(learner-intended-meaning ((he1 x1) (le2 x1 x2) (or1 x2) (sq1 x2) (sm1 x2)))
(learner-utterance (el hexagono a mediana cuadrado azul))
(teacher-classification uninterpretable)
(intended-correction? #f)
(teacher-utterance (el hexagono grande y purpura a la izquierda del cuadrado pequeno y naranja))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...  The learner continues to produce utterances unintelligible
     to the teacher.  In addition to an incorrect general form,
     there seems to be confusion about words for the
     relation le2 and predicates or1, and sm1.

((interaction-number 32)
(situation ((sm1 t1) (gr1 t1) (ci1 t1) (le2 t1 t2) (me1 t2) (bl1 t2) (tr1 t2)))
(learner-intended-meaning ((ci1 x1) (le2 x1 x2) (bl1 x2) (tr1 x2) (me1 x2)))
(learner-utterance (el circulo a azul triangulo mediano))
(teacher-classification uninterpretable)
(intended-correction? #f)
(teacher-utterance (el triangulo azul a la derecha del circulo verde))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...  The learner again uses this incorrect form, but the
     words are more indicative of its intended meaning.

((interaction-number 33)
(situation ((me1 t1) (or1 t1) (st1 t1) (ab2 t1 t2) (bi1 t2) (or1 t2) (ci1 t2)))
(learner-intended-meaning ((ci1 x1) (ab2 x2 x1) (st1 x2)))
(learner-utterance (el circulo debajo del estrella))
(teacher-classification error-in-form)
(intended-correction? #t)
(teacher-utterance (el circulo debajo de la estrella))
(learner-detected-correction? #t)
(new-general-form ())
(learner-understanding ((ci1 x1) (ab2 x2 x1) (st1 x2)))
)

...  The learner uses a correct general form and the
     teacher understands the intended meaning and corrects
```

```
     an error of agreement.  The learner detects the
     correction.  The learner now produces several
     correct utterances, utterances with detected
     corrections of errors in form, and an uninterpretable
     utterance (from the incorrect second general form).

((interaction-number 40)
(situation ((sm1 t1) (bl1 t1) (tr1 t1) (le2 t1 t2) (me1 t2) (pu1 t2) (sq1 t2)))
(learner-intended-meaning ((sq1 x1) (le2 x2 x1) (tr1 x2)))
(learner-utterance (el cuadrado a la derecha del triangulo))
(teacher-classification correct)
(intended-correction? #f)
(teacher-utterance (el cuadrado purpura))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (color1 x1)))
(learner-understanding ((sq1 x1) (pu1 x1)))
)

...  Here the learner acquires a correct sixth general form.

((interaction-number 41)
(situation ((me1 t1) (or1 t1) (sq1 t1) (ab2 t1 t2) (sm1 t2) (bl1 t2) (sq1 t2)))
(learner-intended-meaning ((sq1 x1) (or1 x1)))
(learner-utterance (el cuadrado naranja))
(teacher-classification correct)
(intended-correction? #f)
(teacher-utterance (el cuadrado mediano encima del cuadrado pequeno))
(learner-detected-correction? #f)
(new-general-form ())
(learner-understanding ())
)

...  And proceeds immediately to use it correctly.

((interaction-number 42)
(situation ((bi1 t1) (ye1 t1) (ci1 t1) (le2 t1 t2) (me1 t2) (or1 t2) (st1 t2)))
(learner-intended-meaning ((ci1 x1) (le2 x1 x2) (st1 x2) (or1 x2)))
(learner-utterance (el circulo a la izquierda estrella pequeno))
(teacher-classification error-in-meaning)
(intended-correction? #t)
(teacher-utterance (el circulo a la izquierda de la estrella mediana))
(learner-detected-correction? #t)
(new-general-form ())
(learner-understanding ((ci1 x1) (le2 x1 x2) (st1 x2) (or1 x2)))
)

...  The learner chooses a correct form, but a confusion about
     the meaning of "pequeno" (small, not orange) means that
     the learner's utterance does not express its meaning.
     The teacher detects an error of meaning, and corrects
     "pequeno" (small) to "mediana" (medium), the correct
```

size for the star.  The learner's confusion about the
meaning of "mediana" (medium, not orange) allows it
incorrectly to think that the teacher's utterance has
the same meaning as its own, which results in a
*mistaken* detection of a correction.  This comedy of
errors on the learner's part is followed by several
correct utterances.

```
((interaction-number 45)
(situation ((me1 t1) (bl1 t1) (ci1 t1) (ab2 t1 t2) (sm1 t2) (gr1 t2) (el1 t2)))
(learner-intended-meaning ((ci1 x1) (bl1 x1)))
(learner-utterance (el circulo azul))
(teacher-classification correct)
(intended-correction? #f)
(teacher-utterance (el circulo encima de la elipse verde))
(learner-detected-correction? #f)
(new-general-form ((shape1 x1) (ab2 x1 x2) (shape1 x2) (color1 x2)))
(learner-understanding ((ci1 x1) (ab2 x1 x2) (el1 x2) (gr1 x2)))
)
```

...  A new correct general form is acquired.  This is followed
by several correct utterances and detected corrections
of errors in form.

```
((interaction-number 49)
(situation ((bi1 t1) (or1 t1) (tr1 t1) (ab2 t1 t2) (bi1 t2) (or1 t2) (he1 t2)))
(learner-intended-meaning ((tr1 x1) (ab2 x1 x2) (he1 x2) (or1 x2)))
(learner-utterance (el triangulo encima de la hexagono naranja))
(teacher-classification error-in-form)
(intended-correction? #t)
(teacher-utterance (el triangulo encima del hexagono naranja))
(learner-detected-correction? #t)
(new-general-form ())
(learner-understanding ((tr1 x1) (ab2 x1 x2) (he1 x2) (or1 x2)))
)
```

...  The last interaction of the 50 is followed by the
teacher's summary of its classifications and corrections.
Because the correction probability was 1.0, the
teacher offered a correction for every error in form
and error in meaning.

```
(teacher report(
(correct-utterance 12)
(error-in-meaning 10)
(error-in-form 9)
(uninterpretable 10)
(no-utterance 9)
(intended-correction 19)))
)
```

# B Appendix: Notes on the Sublanguages

## B.1 The 68-Form Languages

The 68-form languages include utterances referring to one object or to two objects with a relation between them. The objects are one of six shapes (square, circle, triangle, star, hexagon, or ellipse), have one of three sizes (big, medium, or small) and have one of six colors (red, orange, yellow, green, blue, or purple.) The relation between them expresses that one object is to the left of, to the right of, above, or below the other object. Each transducer is constructed so that there is exactly one utterance that produces each of the 113, 064 possible expressible meanings, each of which is an instance of one of 68 general forms. Examples of and notes on the 68-form languages in our sample follow. (Note that individual words are represented by ASCII strings.)

**English.**    Examples of teacher utterances: *the purple square above the big yellow hexagon, the big star to the right of the big triangle, the purple triangle to the right of the square, the medium blue hexagon, the red circle.*

**German.**    Examples of teacher utterances: *der orangne kreis, der mittlere stern uber dem grunen kreis, das kleine grune quadrat links vom dem orangnen stern, das kleine blaue hexagon unter dem quadrat, das grosse orangne dreieck links von der ellipse.*

**Greek.**    Examples of teacher utterances: *to exagono, to mikro kokkino tetragono sta aristera tou mesaiou portokali kyklou, i mikri ellipsi, i ellipsi sta deksia tou kyklou, to portokali trigono pano apo to mikro astro.* Phonological rules for articles were not included.

**Hebrew.**    Examples of teacher utterances: *hakochav hagadol hakachol, haelipsa mismol lameshushe, hameshulash hakatan me!al laelipsa haktana hatsehuva, haigul mitachat laelipsa ha!aduma, hameshulash miyamin lameshushe hagadol hasagol.*

**Hungarian.**    Examples of teacher utterances: *a/az piros negyzet, a/az kicsi zold negyzet a/az kozepes csillag folott, a/az piros kor balra a/az negyzettol, a/az csillag a/az kicsi sarga haromszog alatt, a/az kicsi zold negyzet jobbra a/az lila kortol.* The phonological rule to select the article *a* or *az* is omitted, so the article is *a/az* throughout. The grammar chosen has the words for left and right separating the two objects, but the words for above and below following both of them.

**Mandarin.**    Examples of teacher utterances: *da de lu se xing1, liu bian xing zuo bian de huang se zheng fang xing, huang se zheng fang xing, xiao de zheng fang xing xia mian de zhong jian da-xiao de lu se san jiao xing, zhong jian da-xiao de lu se yuan zuo bian de xiao de hong se tuo-yuan.* Because of limitations of our system in dealing with words with more than one meaning and multi-word phrases, we distinguished the senses of *xing* and *xing1* and hyphenated *da-xiao* (medium) and *tuo-yuan* (ellipse).

**Russian.**    Examples of teacher utterances: *bolshoy fioletoviy shestiugolnik, triugolnik pod sinim ellipsom, shestiugolnik c leva ot srednevo ellipsa, kvadrat c prava ot bolshovo shestiugolnika, bolshoy krasniy shestiugolnik c leva ot bolshoy siney zvezdi.* There are three cases of each noun and adjective.

**Spanish.**    Examples of teacher utterances: *el triangulo, el hexagono azul debajo del circulo, la estrella encima de la elipse roja, el cuadrado pequeno y amarillo a la izquierda del cuadrado mediano y naranja, el hexagono grande y azul a la derecha de la elipse grande.* The order of adjectives is fixed as size followed by color.

**Swedish.** Examples of teacher utterances: *cirkeln, den lilla bla fyrkanten, den normala triangeln under den normala oranga stjarnan, den lila ellipsen till hoger om den lilla fyrkanten, den normala roda ellipsen till vanster om hexagonen.*

**Turkish.** Examples of teacher utterances: *kare, mavi karenin ustundeki sari elips, orta yesil cemberin solundaki yildiz, buyuc yildizin altindaki altigen, orta sari cemberin altindaki mavi kare.*

## B.2  The 8-Form Languages

Each 8-form language is a proper sublanguage of the corresponding 68-form language containing all the utterances that refer to one object together with the subset of utterances that refer to two objects and include the size and color of both of them. There are 168 utterances referring to one object and 46,656 utterances that refer to two objects and include the size and color of both, for a total of 46,824 utterances in the 8-form languages. Each meaning is an instance of one of just 8 general forms for each language.