

Experiments Using OSTIA for a Language Production Task

Dana Angluin and Leonor Becerra-Bonache

Department of Computer Science, Yale University

P.O.Box 208285, New Haven, CT, USA

{dana.angluin, leonor.becerra-bonache}@yale.edu

Abstract

The phenomenon of meaning-preserving corrections given by an adult to a child involves several aspects: (1) the child produces an incorrect utterance, which the adult nevertheless understands, (2) the adult produces a correct utterance with the same meaning and (3) the child recognizes the adult utterance as having the same meaning as its previous utterance, and takes that as a signal that its previous utterance is not correct according to the adult grammar. An adequate model of this phenomenon must incorporate utterances and meanings, account for how the child and adult can understand each other's meanings, and model how meaning-preserving corrections interact with the child's increasing mastery of language production. In this paper we are concerned with how a learner who has learned to comprehend utterances might go about learning to produce them.

We consider a model of language comprehension and production based on finite sequential and subsequential transducers. Utterances are modeled as finite sequences of words and meanings as finite sequences of predicates. Comprehension is interpreted as a mapping of utterances to meanings and production as a mapping of meanings to utterances. Previous work (Castellanos et al., 1993; Pieraccini et al., 1993) has applied subsequential transducers and the OSTIA algorithm to the problem of learning to comprehend language; here we apply them to the problem of learning to produce language. For ten natural languages and a limited domain of geometric shapes and their properties and rela-

tions we define sequential transducers to produce pairs consisting of an utterance in that language and its meaning. Using this data we empirically explore the properties of the OSTIA and DD-OSTIA algorithms for the tasks of learning comprehension and production in this domain, to assess whether they may provide a basis for a model of meaning-preserving corrections.

1 Introduction

The role of corrections in language learning has recently received substantial attention in Grammatical Inference. The kinds of corrections considered are mainly *syntactic corrections* based on proximity between strings. For example, a correction of a string may be given by using edit distance (Becerra-Bonache et al., 2007; Kinber, 2008) or based on the shortest extension of the queried string (Becerra-Bonache et al., 2006), among others. In these approaches semantic information is not used.

However, in natural situations, a child's erroneous utterances are corrected by her parents based on the meaning that the child intends to express; typically, the adult's corrections preserve the intended meaning of the child. Adults use corrections in part as a way of making sure they have understood the child's intentions, in order to keep the conversation "on track". Thus the child's utterance and the adult's correction have the same meaning, but the form is different. As Chouinard and Clark point out (2003), 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 something that is not acceptable in the target language. Results in (Chouinard and Clark, 2003) show that adults reformulate erroneous child utterances often enough to help learning. Moreover, these re-

sults show that children can not only detect differences between their own utterance and the adult reformulation, but that they do make use of that information.

Thus in some natural situations, corrections have a *semantic* component that has not been taken into account in previous Grammatical Inference studies. Some interesting questions arise: What are the effects of corrections on learning syntax? Can corrections facilitate the language learning process? One of our long-term goals is to find a formal model that gives an account of this kind of correction and in which we can address these questions. Moreover, such a model might allow us to show that semantic information can simplify the problem of learning formal languages.

A simple computational model of *semantics* and *context* for language learning incorporating semantics was proposed in (Angluin and Becerra-Bonache, 2008). This model accommodates two different tasks: comprehension and production. That paper focused only on the comprehension task and formulated the learning problem as follows. The teacher provides to the learner several example pairs consisting of a situation and an utterance denoting something in the situation; the goal of the learner is to learn the meaning function, allowing the learner to comprehend novel utterances. The results in that paper show that under certain assumptions, a simple algorithm can learn to comprehend an adult’s utterance in the sense of producing *the same sequence of predicates*, even without mastering the adult’s grammar. For example, receiving the utterance *the blue square above the circle*, the learner would be able to produce the sequence of predicates (*bl, sq, ab, ci*).

In this paper we focus on the production task, using sequential and subsequential transducers to model both comprehension and production. Adult production can be modeled as converting a sequence of predicates into an utterance, which can be done with access to the meaning transducer for the adult’s language.

However, we do not assume that the child initially has access to the meaning transducer for the adult’s language; instead we assume that the child’s production progresses through different stages. Initially, child production is modeled as consisting of two different tasks: finding a correct sequence of predicates, and inverting the meaning function to produce a kind of “telegraphic speech”.

For example, from (*gr, tr, le, sq*) the child may produce *green triangle left square*. Our goal is to model how the learner might move from this telegraphic speech to speech that is grammatical in the adult’s sense. Moreover, we would like to find a formal framework in which corrections (in form of expansions, for example, *the green triangle to the left of the square*) can be given to the child during the intermediate stages (before the learner is able to produce grammatically correct utterances) to study their effect on language learning.

We thus propose to model the problem of child language production as a machine translation problem, that is, as the task of translating a sequence of predicate symbols (representing the meaning of an utterance) into a corresponding utterance in a natural language. In this paper we explore the possibility of applying existing automata-theoretic approaches to machine translation to model language production. In Section 2, we describe the use of subsequential transducers for machine translation tasks and review the OSTIA algorithm to learn them (Oncina, 1991). In Section 3, we present our model of how the learner can move from telegraphic to adult speech. In Section 4, we present the results of experiments in the model made using OSTIA. Discussion of these results is presented in Section 5 and ideas for future work are in Section 6.

2 Learning Subsequential Transducers

Subsequential transducers (SSTs) are a formal model of translation widely studied in the literature. SSTs are deterministic finite state models that allow input-output mappings between languages. Each edge of an SST has an associated input symbol and output string. When an input string is accepted, an SST produces an output string that consists of concatenating the output substrings associated with sequence of edges traversed, together with the substring associated with the last state reached by the input string. Several phenomena in natural languages can be easily represented by means of SSTs, for example, the different orders of noun and adjective in Spanish and English (e.g., *un cuadrado rojo - a red square*). Formal and detailed definitions can be found in (Berstel, 1979).

For any SST it is always possible to find an equivalent SST that has the output strings assigned to the edges and states so that they are as close to

the initial state as they can be. This is called an Onward Subsequential Transducer (OST).

It has been proved that SSTs are learnable in the limit from a positive presentation of sentence pairs by an efficient algorithm called OSTIA (Onward Subsequential Transducer Inference Algorithm) (Oncina, 1991). OSTIA takes as input a finite training set of input-output pairs of sentences, and produces as output an OST that generalizes the training pairs. The algorithm proceeds as follows (this description is based on (Oncina, 1998)):

- A prefix tree representation of all the input sentences of the training set is built. Empty strings are assigned as output strings to both the internal nodes and the edges of this tree, and every output sentence of the training set is assigned to the corresponding leaf of the tree. The result is called a *tree subsequential transducer*.
- An *onward tree subsequential transducer* equivalent to the tree subsequential transducer is constructed by moving the longest common prefixes of the output strings, level by level, from the leaves of the tree towards the root.
- Starting from the root, all pairs of states of the onward tree subsequential transducer are considered in order, level by level, and are merged if possible (i.e., if the resulting transducer is subsequential and does not contradict any pair in the training set).

SSTs and OSTIA have been successfully applied to different translation tasks: Roman numerals to their decimal representations, numbers written in English to their Spanish spelling (Oncina, 1991) and Spanish sentences describing simple visual scenes to corresponding English and German sentences (Castellanos et al., 1994). They have also been applied to language understanding tasks (Castellanos et al., 1993; Pieraccini et al., 1993).

Moreover, several extensions of OSTIA have been introduced. For example, OSTIA-DR incorporates domain (input) and range (output) models in the learning process, allowing the algorithm to learn SSTs that accept only sentences compatible with the input model and produce only sentences compatible with the output model (Oncina

and Varo, 1996). Experiments with a language understanding task gave better results with OSTIA-DR than with OSTIA (Castellanos et al., 1993). Another extension is DD-OSTIA (Oncina, 1998), which instead of considering a lexicographic order to merge states, uses a heuristic order based on a measure of the equivalence of the states. Experiments in (Oncina, 1998) show that better results can be obtained by using DD-OSTIA in certain translation tasks from Spanish to English.

3 From telegraphic to adult speech

To model how the learner can move from telegraphic speech to adult speech, we reduce this problem to a translation problem, in which the learner has to learn a mapping from sequences of predicates to utterances. As we have seen in the previous section, SSTs are an interesting approach to machine translation. Therefore, we explore the possibility of modeling language production using SSTs and OSTIA, to see whether they may provide a good framework to model corrections.

As described in (Angluin and Becerra-Bonache, 2008), after learning the meaning function the learner is able to assign correct meanings to utterances, and therefore, given a situation and an utterance that denotes something in the situation, the learner is able to point correctly to the object denoted by the utterance. To simplify the task we consider, we make two assumptions about the learner at the start of the production phase: (1) the learner's lexicon represents a correct meaning function and (2) the learner can generate correct sequences of predicates.

Therefore, in the initial stage of the production phase, the learner is able to produce a kind of "telegraphic speech" by inverting the lexicon constructed during the comprehension stage. For example, if the sequence of predicates is (*bl, sq, ler, ci*), and in the lexicon *blue* is mapped to *bl, square* to *sq, right* to *ler* and *circle* to *ci*, then by inverting this mapping, the learner would produce *blue square right circle*.

In order to explore the capability of SSTs and OSTIA to model the next stage of language production (from telegraphic to adult speech), we take the training set to be input-output pairs each of which contains as input a sequence of predicates (e.g., (*bl, sq, ler, ci*)) and as output the corresponding utterance in a natural language (e.g., *the blue square to the right of the circle*). In this example,

the learner must learn to include appropriate function words. In other languages, the learner may have to learn a correct choice of words determined by gender, case or other factors. (Note that we are not yet in a position to consider corrections.)

4 Experiments

Our experiments were made for a limited domain of geometric shapes and their properties and relations. This domain is a simplification of the *Miniature Language Acquisition* task proposed by Feldman et al. (Feldman et al., 1990). Previous applications of OSTIA to language understanding and machine translation have also used adaptations and extensions of the Feldman task.

In our experiments, we have predicates for three different shapes (circle (*ci*), square (*sq*) and triangle (*tr*)), three different colors (blue (*bl*), green (*gr*) and red (*re*)) and three different relations (to the left of (*le*), to the right of (*ler*), and above (*ab*)). We consider ten different natural languages: Arabic, English, Greek, Hebrew, Hindi, Hungarian, Mandarin, Russian, Spanish and Turkish.

We created a data sequence of input-output pairs, each consisting of a predicate sequence and a natural language utterance. For example, one pair for Spanish is $((ci, re, ler, tr), el\ círculo\ rojo\ a\ la\ derecha\ del\ triangulo)$. We ran OSTIA on initial segments of the sequence of pairs, of lengths 10, 20, 30, . . . , to produce a sequence of subsequential transducers. The whole data sequence was used to test the correctness of the transducers generated during the process. An error is counted whenever given a data pair (x, y) , the subsequential transducer translates x to y' , and $y' \neq y$. We say that OSTIA has *converged* to a correct transducer if all the transducers produced afterwards have the same number of states and edges, and 0 errors on the whole data sequence.

To generate the sequences of input-output pairs, for each language we constructed a meaning transducer capable of producing the 444 different possible meanings involving one or two objects. We randomly generated 400 unique (non-repeated) input-output pairs for each language. This process was repeated 10 times. In addition, to investigate the effect of the order of presentation of the input-output pairs, we repeated the data generation process for each language, sorting the pairs according to a length-lex ordering of the utterances.

We give some examples to illustrate the trans-

ducers produced. Figure 1 shows an example of a transducer produced by OSTIA after just ten pairs of input-output examples for Spanish. This transducer correctly translates the ten predicate sequences used to construct it, but the data is not sufficient for OSTIA to generalize correctly in all cases, and many other correct meanings are still incorrectly translated. For example, the sequence (ci, bl) is translated as *el círculo a la izquierda del círculo verde azul* instead of *el círculo azul*.

The transducers produced after convergence by OSTIA and DD-OSTIA correctly translate all 444 possible correct meanings. Examples for Spanish are shown in Figure 2 (OSTIA) and Figure 3 (DD-OSTIA). Note that although they correctly translate all 444 correct meanings, the behavior of these two transducers on other (incorrect) predicate sequences is different, for example on (tr, tr) .

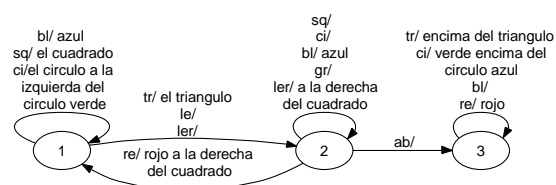


Figure 1: Production task, OSTIA. A transducer produced using 10 random unique input-output pairs (predicate sequence, utterance) for Spanish.

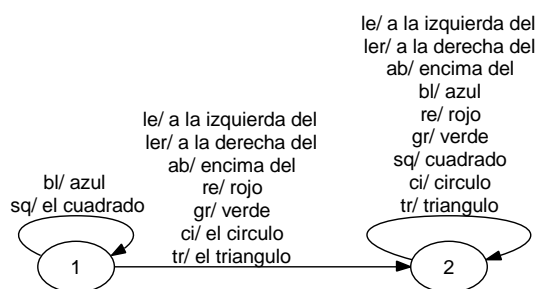


Figure 2: Production task, OSTIA. A transducer produced (after convergence) by using random unique input-output pairs (predicate sequence, utterance) for Spanish.

Different languages required very different numbers of data pairs to converge. Statistics on the number of pairs needed until convergence for OSTIA for all ten languages for both random unique and random unique sorted data sequences are shown in Table 1. Because better results were reported using DD-OSTIA in machine translation

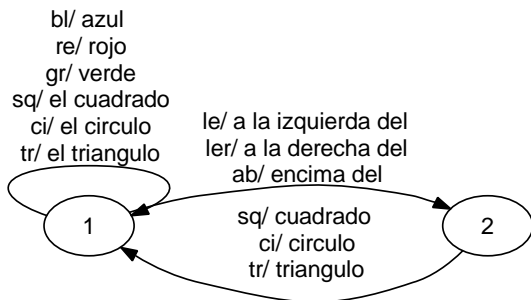


Figure 3: Production task, DD-OSTIA. A transducer produced (after convergence) using random unique input-output pairs (predicate-sequence, utterance) for Spanish.

Language	# Pairs	# Sorted Pairs
Arabic	150	200
English	200	235
Greek	375	400
Hebrew	195	30
Hindi	380	350
Hungarian	365	395
Mandarin	45	150
Russian	270	210
Spanish	190	150
Turkish	185	80

Table 1: Production task, OSTIA. The entries give the median number of input-output pairs until convergence in 10 runs. For Greek, Hindi and Hungarian, the median for the unsorted case is calculated using all 444 random unique pairs, instead of 400.

tasks (Oncina, 1998), we also tried using DD-OSTIA for learning to translate a sequence of predicates to an utterance. We used the same sequences of input-output pairs as in the previous experiment. The results obtained are shown in Table 2.

We also report the sizes of the transducers learned by OSTIA and DD-OSTIA. Table 3 and Table 4 show the numbers of states and edges of the transducers after convergence for each language. In case of disagreements, the number reported is the mode.

To answer the question of whether production is harder than comprehension in this setting, we also considered the *comprehension task*, that is, to translate an utterance in a natural language into the corresponding sequence of predicates.

Language	# Pairs	# Sorted Pairs
Arabic	80	140
English	85	180
Greek	350	400
Hebrew	65	80
Hindi	175	120
Hungarian	245	140
Mandarin	40	150
Russian	185	210
Spanish	80	150
Turkish	50	40

Table 2: Production task, DD-OSTIA. The entries give the median number of input-output pairs until convergence in 10 runs. For Greek, Hindi and Hungarian, the median for the unsorted case is calculated using all 444 random unique pairs, instead of 400.

Languages	#states	#edges
Arabic	2	20
English	2	20
Greek	9	65
Hebrew	2	20
Hindi	7	58
Hungarian	3	20
Mandarin	1	10
Russian	3	30
Spanish	2	20
Turkish	4	31

Table 3: Production task, OSTIA. Sizes of transducers at convergence.

The comprehension task was studied by Oncina et al. (Castellanos et al., 1993). They used English sentences, with a more complex version of the Feldman task domain and more complex semantic representations than we use. Our results are presented in Table 5. The number of states and edges of the transducers after convergence is shown in Table 6.

5 Discussion

It should be noted that because the transducers output by OSTIA and DD-OSTIA correctly reproduce all the pairs used to construct them, once either algorithm has seen all 444 possible data pairs in either the production or the comprehension task, the resulting transducers will correctly translate all correct inputs. However, state-merging in the al-

Languages	#states	#edges
Arabic	2	17
English	2	16
Greek	9	45
Hebrew	2	13
Hindi	7	40
Hungarian	3	20
Mandarin	1	10
Russian	3	23
Spanish	2	13
Turkish	3	18

Table 4: Production task, DD-OSTIA. Sizes of transducers at convergence.

Languages	OSTIA	DD-OSTIA
Arabic	65	65
English	60	20
Greek	325	60
Hebrew	90	45
Hindi	60	35
Hungarian	40	45
Mandarin	60	40
Russian	280	55
Spanish	45	30
Turkish	60	35

Table 5: Comprehension task, OSTIA and DD-OSTIA. Median number (in 10 runs) of input-output pairs until convergence using a sequence of 400 random unique pairs of (utterance, predicate sequence).

gorithms induces compression and generalization, and the interesting questions are how much data is required to achieve correct generalization, and how that quantity scales with the complexity of the task. This are very difficult questions to approach analytically, but empirical results can offer valuable insights.

Considering the comprehension task (Tables 5 and 6), we see that OSTIA generalizes correctly from at most 15% of all 444 possible pairs except in the cases of Greek, Hebrew and Russian. DD-OSTIA improves the OSTIA results, in some cases dramatically, for all languages except Hungarian. DD-OSTIA achieves correct generalization from at most 15% of all possible pairs for all ten languages. Because the meaning function for all ten language transducers is independent of the state, in each case there is a 1-state sequential trans-

Languages	#states	#edges
Arabic	1	15
English	1	13
Greek	2	25
Hebrew	1	13
Hindi	1	13
Hungarian	1	14
Mandarin	1	17
Russian	1	24
Spanish	1	14
Turkish	1	13

Table 6: Comprehension task, OSTIA and DD-OSTIA. Sizes of transducers at convergence using 400 random unique input-output pairs (utterance, predicate sequence). In cases of disagreement, the number reported is the mode.

ducer that achieves correct translation of correct utterances into predicate sequences. OSTIA and DD-OSTIA converged to 1-state transducers for all languages except Greek, for which they converged to 2-state transducers. Examining one such transducer for Greek, we found that the requirement that the transducer be “onward” necessitated two states. These results are broadly compatible with the results obtained by Oncina et al. (Castellanos et al., 1993) on language understanding; the more complex tasks they consider also give evidence that this approach may scale well for the comprehension task.

Turning to the production task (Tables 1, 2, 3 and 4), we see that providing the random samples with a length-lex ordering of utterances has inconsistent effects for both OSTIA and DD-OSTIA, sometimes dramatically increasing or decreasing the number of samples required. We do not further consider the sorted samples.

Comparing the production task with the comprehension task for OSTIA, the production task generally requires substantially more random unique samples than the comprehension task for the same language. The exceptions are Mandarin (production: 45 and comprehension: 60) and Russian (production: 270 and comprehension: 280). For DD-OSTIA the results are similar, with the sole exception of Mandarin (production: 40 and comprehension: 40). For the production task DD-OSTIA requires fewer (sometimes dramatically fewer) samples to converge than OSTIA. However, even with DD-OSTIA the number of sam-

ples is in several cases (Greek, Hindi, Hungarian and Russian) a rather large fraction (40% or more) of all 444 possible pairs. Further experimentation and analysis is required to determine how these results will scale.

Looking at the sizes of the transducers learned by OSTIA and DD-OSTIA in the production task, we see that the numbers of states agree for all languages except Turkish. (Recall from our discussion in Section 4 that there may be differences in the behavior of the transducers learned by OSTIA and DD-OSTIA at convergence.) For the production task, Mandarin gives the smallest transducer; for this fragment of the language, the translation of correct predicate sequences into utterances can be achieved with a 1-state transducer. In contrast, English and Spanish both require 2 states to handle articles correctly. For example, in the transducer in Figure 3, the predicate for a circle (*ci*) is translated as *el círculo* if it occurs as the first object (in state 1) and as *círculo* if it occurs as second object (in state 2) because *del* has been supplied by the translation of the intervening binary relation (*le*, *ler*, or *ab*.) Greek gives the largest transducer for the production task, with 9 states, and requires the largest number of samples for DD-OSTIA to achieve convergence, and one of the largest numbers of samples for OSTIA. Despite the evidence of the extremes of Mandarin and Greek, the relation between the size of the transducer for a language and the number of samples required to converge to it is not monotonic.

In our model, one reason that learning the production task may in general be more difficult than learning the comprehension task is that while the mapping of a word to a predicate does not depend on context, the mapping of a predicate to a word or words does (except in the case of our Mandarin transducer.) As an example, in the comprehension task the Russian words *triugolnik*, *triugolnika* and *triugonikom* are each mapped to the predicate *tr*, but the reverse mapping must be sensitive to the context of the occurrence of *tr*.

These results suggest that OSTIA or DD-OSTIA may be an effective method to learn to translate sequences of predicates into natural language utterances in our domain. However, some of our objectives seem incompatible with the properties of OSTIA. In particular, it is not clear how to incorporate the learner’s initial knowledge of the lexicon and ability to produce “telegraphic

speech” by inverting the lexicon. Also, the intermediate results of the learning process do not seem to have the properties we expect of a learner who is progressing towards mastery of production. That is, the intermediate transducers perfectly translate the predicate sequences used to construct them, but typically produce other translations that the learner (using the lexicon) would know to be incorrect. For example, the intermediate transducer from Figure 1 translates the predicate sequence (*ci*) as *el círculo a la izquierda del círculo verde*, which the learner’s lexicon indicates should be translated as (*ci*, *le*, *ci*, *gr*).

6 Future work

Further experiments and analysis are required to understand how these results will scale with larger domains and languages. In this connection, it may be interesting to try the experiments of (Castellanos et al., 1993) in the reverse (production) direction. Finding a way to incorporate the learner’s initial lexicon seems important. Perhaps by incorporating the learner’s knowledge of the input domain (the legal sequences of predicates) and using the domain-aware version, OSTIA-D, the intermediate results in the learning process would be more compatible with our modeling objectives. Coping with errors will be necessary; perhaps an explicitly statistical framework for machine translation should be considered.

If we can find an appropriate model of how the learner’s language production process might evolve, then we will be in a position to model meaning-preserving corrections. That is, the learner chooses a sequence of predicates and maps it to a (flawed) utterance. Despite its flaws, the learner’s utterance is understood by the teacher (i.e., the teacher is able to map it to the sequence of predicates chosen by the learner) and responds with a correction, that is, a correct utterance for that meaning. The learner, recognizing that the teacher’s utterance has the same meaning but a different form, then uses the correct utterance (as well as the meaning and the incorrect utterance) to improve the mapping of sequences of predicates to utterances.

It is clear that in this model, corrections are not *necessary* to the process of learning comprehension and production; once the learner has a correct lexicon, the utterances of the teacher can be translated into sequences of predicates, and the pairs

of (predicate sequence, utterance) can be used to learn (via an appropriate variant of OSTIA) a perfect production mapping. However, it seems very likely that corrections can make the process of learning a production mapping easier or faster, and finding a model in which such phenomena can be studied remains an important goal of this work.

7 Acknowledgments

The authors sincerely thank Prof. Jose Oncina for the use of his programs for OSTIA and DD-OSTIA, as well as his helpful and generous advice. The research of Leonor Becerra-Bonache was supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme.

References

- Dana Angluin and Leonor Becerra-Bonache. 2008. Learning Meaning Before Syntax. *ICGI*, 281–292.
- Leonor Becerra-Bonache, Colin de la Higuera, J.C. Janodet, and Frederic Tantini. 2007. Learning Balls of Strings with Correction Queries. *ECML*, 18–29.
- Leonor Becerra-Bonache, Adrian H. Dediu, and Cristina Tirnauca. 2006. Learning DFA from Correction and Equivalence Queries. *ICGI*, 281–292.
- Jean Berstel. 1979. Transductions and Context-Free Languages. *PhD Thesis*, Teubner, Stuttgart, 1979.
- Antonio Castellanos, Enrique Vidal, and Jose Oncina. 1993. Language Understanding and Subsequential Transducers. *ICGI*, 11/1–11/10.
- Antonio Castellanos, Ismael Galiano, and Enrique Vidal. 1994. Applications of OSTIA to machine translation tasks. *ICGI*, 93–105.
- Michelle M. Chouinard and Eve V. Clark. 2003. Adult Reformulations of Child Errors as Negative Evidence. *Journal of Child Language*, 30:637–669.
- Jerome A. Feldman, George Lakoff, Andreas Stolcke, and Susan Hollback Weber. 1990. Miniature Language Acquisition: A touchstone for cognitive science. Technical Report, TR-90-009. International Computer Science Institute, Berkeley, California. April, 1990.
- Efim Kinber. 2008. On Learning Regular Expressions and Patterns Via Membership and Correction Queries. *ICGI*, 125–138.
- Jose Oncina. 1991. Aprendizaje de lenguajes regulares y transducciones subsecuenciales. *PhD Thesis*, Universitat Politecnica de Valencia, Valencia, Spain, 1998.
- Jose Oncina. 1998. The data driven approach applied to the OSTIA algorithm. *ICGI*, 50–56.
- Jose Oncina and Miguel Angel Varo. 1996. Using domain information during the learning of a subsequential transducer. *ICGI*, 301–312.
- Roberto Pieraccini, Esther Levin, and Enrique Vidal. 1993. Learning how to understand language. *EuroSpeech'93*, 448–458.