

# Creating AI: A unique interplay between the development of learning algorithms and their education

Anat Treister-Goren and Jason L. Hutchens

## Abstract

**Artificial Intelligence Enterprises is a new research and development company devoted to the creation of Artificial Intelligence (specifically, the development of computer programs featuring human-like conversational capabilities). As such, we believe that our technology will be the first to pass the Turing Test. The philosophy driving this project is the firm conviction that returning to Turing's original vision of building a child machine and then training it to perform is the only way to achieve our goal. This paper introduces our development cycle, which is based on a continuous interplay between developing learning algorithms and training them to converse, and gives some encouraging examples of the performance of a system developed in this environment.**

## 1 Introduction

In 1950 Alan Turing considered the question "Can machines think?" Turing's answer to this question was to define the meaning of the term 'think' in terms of a conversational scenario, whereby if an interrogator cannot reliably distinguish between a machine and a human based solely on their conversational ability, then the machine could be said to be thinking<sup>1</sup>. This procedure is nowadays referred to as the Turing Test.

The field of artificial intelligence has largely ignored Turing's strict criterion. Today it encompasses topics as diverse as intelligent agents, chatterbots, pattern recognition systems, voice recognition systems and expert systems with applications in medicine, finance, entertainment, business, and manufacturing. Even though important work has been conducted in terms of the sophistication and expertise of these systems, the vision that motivated the birth of the field has not yet been fulfilled: there is neither cooperation nor agreement amongst its researchers. The

unfortunate result of this trend is that true advancement is inhibited.

Artificial Intelligence Enterprises (*AI*) is a newly formed company, which is dedicated to a decade-long research project aimed at creating a computer program capable of passing the Turing Test. This is the first commercially oriented research and development effort, with complete financial backing, that is committed to the creation of 'true' artificial intelligence: computer programs with human-like, unrestricted, and natural conversational capability.

The methods employed to strive towards this ambitious goal make *AI*'s project unique. General learning algorithms are researched and developed to improve the performance of various language tasks. The resulting system is then trained in a black-box fashion: knowledge of its inner workings is not allowed to influence the training strategy, although a particular training strategy will often evolve to suit a particular implementation. Specific language performance milestones are aimed at during training, and evaluation is performed with reference to human lingual development. Ongoing feedback of training results is proving crucial to the efficient development of the learning algorithms.

In this paper we shall describe the research philosophy being applied at *AI*, while refraining from providing detailed results of our work and from discussing technical issues at length. We begin by describing the Turing Test and by sharing our opinion as to why computer programs designed to hold conversations in natural language have hitherto failed to pass the test. This leads us, via a discussion on behaviorism, into an overview of our research and development work. We follow this with a discussion of our training and evaluation strategies, showing how the interplay between the two proves beneficial. We conclude with some encouraging behavior exhibited by our system during the initial steps of the training process.

## 2 Looking Backwards

### 2.1 The Turing Test

The Turing Test is an appealing measure of artificial intelligence because, as Turing himself writes, it "has the advantage of drawing a fairly sharp line between the physical and the intellectual capacities of a man". The Loebner Contest, held annually since 1991, is an instantiation of the Turing

Test<sup>2</sup>. The sophistication and performance of computer programs entered into the contest, or lack thereof, bears out our introductory remark that the field of artificial intelligence has largely ignored the Turing Test. In a recent thorough review of conversational systems, Hasida and Den emphasize the absurdity of performance in the Loebner Contest<sup>3</sup>. They assert that since the Turing Test requires that systems “talk like people”, and since no system currently meets this requirement, the ad-hoc techniques which the Loebner Contest subsequently encourages make little contribution to the advancement of dialog technology.

We believe that the Turing Test is an appropriate evaluation criterion for the perception of intelligence, and therefore our approach makes the assumption that intelligence is manifested in conversational skills. We firmly believe that engaging in domain-unrestricted conversation is the most critical evidence of intelligence.

### 2.1.1 Turing's Child Machine

Turing concluded his classic paper by theorizing on the design of a computer program, which would be capable of passing the Turing Test. He correctly anticipated the difficulties in simulating adult level conversation, and proposed, “instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain.”<sup>1</sup>

Turing regarded language as an acquired skill and recognized the importance of avoiding the hard wiring of the computer program wherever possible. He viewed language learning in a behavioristic light and believed that the language channel, narrow though it may be, is sufficient to transmit the information that the child machine requires in order to acquire language.

### 2.2 The Traditional Approach

The traditional approach to conversational system design has been to treat language as a knowledge base, and to hard-wire the rules of this knowledge base to generate conversations. This approach has failed to produce anything more sophisticated than domain-restricted dialog systems, which lack the kind of flexibility, openness, and capacity to learn that are the very essence of human intelligence.

Contrary to Turing's prediction that at the turn of the millennium computer programs will participate in the Turing Test so effectively that an average interrogator will have no more than a seventy percent chance of making the right identification after five minutes of questioning, no true conversational systems have yet been produced, and none have passed an unrestricted Turing Test. This may be due in part to the unfortunate fact that Turing's idea of the child machine has remained unexplored.

The failure to generate conversational capability is most likely related to some of the changes that took place since the 1950's in the field of child language research and linguistics in general. A revolution inspired by Chomsky's transformational grammar<sup>4</sup> occurred, dictating the implementation of hard-wired rules to generate language. The Chomskian revolution pushed aside the competing behaviorist theory of language headed by Skinner<sup>5</sup>. Computational implementations based on the Chomskian philosophy became the standard method for trying to generate conversational capability, yielding disappointing results. It is our thesis that true conversational abilities are more easily obtainable via the currently neglected behavioristic approach.

### 2.3 Verbal Behavior

Behaviorism focuses on the observable and measurable aspects of behavior and the search for observable environmental conditions, known as stimuli, that co-occur with and predict the appearance of specific behavior, known as responses<sup>6</sup>. Behaviorists do not deny the existence of internal mechanisms: they do recognize that studying the physiological basis is necessary for a better understanding of behavior. What behaviorists object to are internal structures or processes with no specific physical correlate inferred from behavior. Therefore, they object to the kind of grammatical structures proposed by linguists (particularly Chomskian ones), claiming that these only complicate explanations of language acquisition<sup>7</sup>. They favor a functional rather than a structural approach, with a focus on the stimuli that evokes verbal behavior, and the consequences in language performance. We believe this to be the right approach for the generation of artificial intelligence.

Skinner argues that psycholinguists should ignore traditional categories of linguistic units and should instead treat language as they would any other behavior. That is, since

language is regarded as a skill that is not essentially different from any other behavior, generating and understanding language must therefore be controlled by stimuli from the environment in the form of reinforcement, imitation, and successive approximations to mature performance.

### 3 The Ai Approach: Research and Development

Nobody really understands how the human brain works, nor do we fully grasp the process by which human beings acquire and use natural language. Language is a complicated artifact, and it is impossible for us to observe the low-level processes, which give rise to it. It is imperative, therefore, that we refrain from hardwiring any *a priori* rules into the system. Any other approach would pollute the system with inevitable misconceptions and hinder its development.

We apply the behavioristic framework to a general learning mechanism, with the goal of having it acquire natural language and use it conversationally, via an iterative development-training-development cycle. Indeed, it is our belief that basic information processing mechanisms enable the human brain to handle language, and success at transferring learning algorithms developed for image recognition to the language domain, lends weight to this argument.

The development cycle employed at *Ai* focuses on the progressive specialization of general learning algorithms to the problem of language acquisition. The system consists of a set of learning capabilities coupled with a drive to perform. The development of the learning capabilities, which should be as simple and as general as possible, is driven by the demand for performance in natural language conversation defined by the system's trainer.

#### 3.1 The Nature of Learning

Learning in general is intimately entwined with the acts of prediction and compression. Every living thing constantly makes predictions about the world around it. Will the approaching fanged, never-before-seen creature pounce? Does glimpsing one berry on the forest floor mean others may be found nearby? If observed behavior is mimicked, will the observed rewards be attained? Should I tell Granny that I missed what she just said, or should I fly by the seat of my pants and reply to what I think I heard?

Being able to predict well is conditional on one's ability to draw conclusions from one's experience in order to react to a novel event. However, so much of our experience counts for nothing, and determining the important features of one's history to make a quick and accurate appraisal of the present and a useful prediction of the future is the essence of intelligence. Learning (and, particularly, learning language) may therefore be seen as an act of efficiently compressing the past. We remove redundancy and, like searching for a needle in a haystack, hone in on the aspects of our experience that are most useful to the situation at hand.

#### 3.2 System Architecture

We may consider our system as a black box whose environment consists of a symbolic time series (the sequence of symbols, representing natural language utterances, given to the system as input and generated by the system as output) and which experiences feedback from the environment (positive or negative reinforcement administered by the trainer).

Our system treats each observed symbol, which, to begin with, is a single ASCII character, as a stimulus for its successor. Each symbol in the alphabet known to the system therefore functions as a predictor for the symbols, which may follow it, and we find it beneficial to make these predictors stochastic—their predictions are expressed in the form of a probability distribution over the alphabet.

The predictors update their probability estimates on the basis of observed symbol sequences, and are used to generate novel symbol sequences. Claude Shannon, the father of Information Theory, was doing as much over fifty years ago<sup>8</sup>. We therefore have a system which can learn, in a fashion, from its previous experience, and which can generate a sequence of symbols that satisfies the constraints it knows of.

#### 3.3 Reinforced Learning

The trainer of the system may administer reinforcement by accepting the generated sequence of symbols, rejecting it, or accepting the sequence up to a point and rejecting the generation from then on. The rejection of a symbol is evidence to the system that the symbol used as stimulus (the contents of the system's short-term memory, in effect) failed to capture sufficient contextual information to

make a good prediction about which symbol should follow it.

Failure to perform is an invitation to learn. As in real life, the system improves its performance by learning from its mistakes. The main thrust of our work is to develop learning algorithms which are as general as possible, and to let the system itself decide which kind of learning is most advantageous to it in a particular situation. As an example, we shall briefly describe two different kinds of meta-structure: sub-objects and quotient-objects.

### 3.3.1 Sub-Objects

Sub-objects are sequences of symbols, and our implementation restricts sub-objects to symbol pairs for reasons, which will soon be made apparent. Creating a new symbol from a pair of existing symbols gives the system more context, which may allow it to make a more accurate prediction. Our system begins on the level of ASCII characters but can very quickly learn English words merely by forming sub-objects from symbol pairs.

We find that many of the symbols formed by the system only have a limited lifetime. They are useful for a while, but are rapidly superseded by newer symbols. They act as scaffolding, allowing the system to learn useful, complicated structures without requiring the hard wiring of complicated learning procedures.

Sub-objects may be formed by applying various information theoretic measures. We can calculate the information contained in one symbol in the context of its predecessor and form a sub-object when the information is low. We can calculate the entropy<sup>i</sup> of a symbol (remember that all symbols are predictors and therefore have an entropy associated with them) and form a sub-object from the symbol and its predecessor when the entropy is high. We can measure the correlation between two symbols via the point wise mutual information and join them together to form a sub-object when they are strongly correlated. Alternatively, we can hypothesize all possible sub-objects and commit to a hypothesis whenever doing so would be beneficial to the system.

---

<sup>i</sup> Entropy is an information theoretic measure of uncertainty associated with a probability distribution.

### 3.3.2 Quotient-Objects

Researchers in data-driven learning systems are all too familiar with the problems caused by paucity of the data. The system is constantly faced with situations where simply not enough data has been observed to guarantee an accurate prediction. The formation of quotient-objects, which are equivalence classes of symbols, enables the system to transfer its knowledge of one symbol, which it may have observed quite often, to another symbol, which may occur quite rarely. As with sub-objects, our implementation restricts quotient-objects to classes that contain exactly two symbols. However, since each quotient-object formed becomes a new symbol in its own right, we find that large classes are rapidly formed, and these may be hierarchically decomposed due to the nature of their formation.

### 3.3.3 Making Mistakes

The fact that the system is still rather shortsighted often results in the formation of new symbols which are locally useful but which may hinder performance on a wider scale. A new symbol may only be useful in certain contexts, for example, but the system is blind to this, as it cannot assess the effect of a new symbol without forming it first, and by then it is too late. The system therefore needs to be given the ability to reverse bad decisions it has made in the past, and this is, in effect, a third kind of learning. Such implementations are made trivial by the fact that all sub-objects and quotient-objects consist of two lower-level symbols only. Decomposing a sub-object is a simple task of cleaving it in two, while decomposing a quotient-object is merely a case of replacing it with the symbol actually observed in the data. This process of correcting generalizations in certain contexts facilitates the formation of context-dependent structures.

### 3.3.4 Justification

Sub-objects and quotient-objects are, in a certain sense, orthogonal. What we gain by forming one kind of structure we lose with the other. The payoff between discrimination and generalization seems to be an important factor in learning systems, and gives rise to seemingly creative generations while satisfying contextual constraints.

The advantage of our approach is that the system does not distinguish between symbols at various levels of abstraction. All symbols are predictors, and the only symbols, which

exist, are those, which have proven themselves to be useful. The learning process is a process of evolution: new symbols are born, and compete to perform with existing symbols. Symbols live and die as a result of reinforcement administered by a human trainer.

Although we cannot disclose the details of our learning algorithms, it should be noted that our approach is neither top-down nor bottom-up, but a compromise between the two. We begin from a completely bottom-up perspective, but the formation of new symbols quickly results in a limited top-down view of the data being made available to the system—it can see further than it could in the past, and this information can potentially be used to guide the formation of new symbols.

### 3.4 Generation

Predictors may be used generatively simply by emitting a symbol at random in accordance with the probability distribution given by the system as its prediction. Although simple, this behavior is not sufficient to ensure that the generations are relevant in the context of a natural language dialogue. A more sophisticated process is necessary.

The system needs to be given a drive to perform, and the most obvious drive is directly related to the act of reinforcement. The system should ‘enjoy’ receiving positive reinforcement, and should ‘dislike’ receiving negative reinforcement. It is a simple matter to combine the sequence of predictions made by the system with a model, which estimates the likelihood of receiving punishments and rewards in various contexts, allowing the system as a whole to make the generation, which maximizes its chance of “experiencing pleasure”.

### 3.5 Feedback Loop

We have described the architecture of a basic learning system designed to acquire natural language. Needless to say, we have merely outlined some initial steps and touched upon the principles underlying our research. Our novel approach is based on a feedback loop between the processes of researching and developing new learning algorithms and teaching and evaluating each such implementation.

The research and development team defines the requirements based on reports received from the training and evaluation team. This feedback shapes the direction of the research.

Feedback consists of the abilities which the system has demonstrated, the areas in which it is lacking, the observed behavior, and the desired behavior. Learning algorithms are then developed to produce the desired behavior without compromising the generality of the system. We have found that the training and evaluation phase often gives surprising results, with the system exhibiting unexpected behavior, and with the learning abilities of the system being applied in “creative” ways.

## 4 The Ai Approach: Training and Evaluation

This part of the paper will introduce our training methodology, which combines the behavioral and developmental principles into a solid method of teaching the system to converse. The section first discusses the developmental model and explains the subjective and objective components before touching upon some critical problems in assessing current conversational systems.

### 4.1 The Developmental Principle

We propose that the proper training model must be based on a developmental principle. The developmental model dictates a chronological step-by-step process for creating conversational capability. The child language acquisition field has developed descriptive milestones composed of typical language performance descriptors for the different developmental stages<sup>9</sup>. Extensive analysis tools are available for analyzing conversational performance to determine lingual maturity<sup>10</sup>.

Using the developmental principle validates the Turing Test according to which the judgment of intelligence is in the eye of the beholder. Thus, human perception of intelligence is always influenced by the expectation level of the judge as regards the person or entity under scrutiny (obviously, intelligence in monkeys, children, or university professors will be judged differently). The initial evaluation of maturity level will set up the right expectation level for a valid judgment of conversational capability or intelligence. Once developmental evaluations of systems become the standard procedures, subjective judgments of systems’ intelligence will become valid.

This presents an additional advantage to using the developmental model: a standard evaluation procedure will determine conversational level across various systems. Programs can then be compared as to being at

the level of toddlers, children, adolescents, or adults in terms of conversational capability. Moreover, this approach enables evaluation not only across programs but also within a given program.

#### 4.1.1 Success in Other Fields

Developmental principles have enabled evaluation and treatment programs in other fields formerly suffering from a lack of organizational and evaluative principles<sup>11</sup> and have been especially useful in areas that border on the question of intelligence<sup>12</sup>. Normative developmental language data have enabled the establishment of diagnostic scales, evaluation criteria, and treatment programs for developmentally delayed populations. The developmental approach has proven to be a powerful tool in other areas, such as schizophrenic thought disorder<sup>13</sup>. Clinicians often found themselves unable to capture the communicative problem of patients in order to assess their intelligence level or cognitive capability, let alone to decipher medication treatment effects.

#### 4.2 The Evaluation Process

The ability to converse is complex, continuous, and incremental in nature. Thus we propose to complement our subjective impression of intelligence with objective incremental metrics that will be used both as our guidelines when we train to converse and as evaluation standards.

##### 4.2.1 Objective Parameters

The objective parameters consist of individual metrics<sup>ii</sup> that capture specific aspects of the lingual performance and collectively provide a complete behavioral description of the child's language development stage. Numerous researchers describe human language development using these metrics to analyze transcribed conversations between children and their caretakers. Examples of some lingual developmental metrics, which increase quantitatively with age, are vocabulary size (the number of different words spoken), mean length of utterance (the mean number of morphemes spoken per utterance), syntactic complexity (the ability to use embedding to connect sentences) and the use of pronominal and referential forms.

---

<sup>ii</sup> We use the term "metric" in its non-mathematical sense of relating to measurement.

#### 4.3 Another Feedback Loop

The training process is driven by the need to achieve certain performance milestones. These milestones dictate the kind of reinforcement that trainers give to the conversing program. That is, in order to achieve a certain performance, the program will be reinforced to produce specific lingual behaviors over others, those that are expected for the trained conversational level. The feedback to the algorithm department includes reporting success on language tasks and limitations on other tasks. Being language researchers, trainers also provide feedback in the form of assumptions on missing or required human learning abilities that children are thought to have. In summary, the detailed description of the program's conversational performance enables the development team to understand its capabilities, as they are instantiated in natural language. This performance report and analysis reveals aspects that need to be developed which would not otherwise be revealed.

### 5 Preliminary results

Currently we have evolved through two development cycles or two back-and-forth interactions between the algorithmic department and the training team. In the first cycle, we established a base age reflecting the system's capabilities. This base age was determined using both the subjective and objective assessments as described below. The system is internally code-named "Hal".

#### 5.1 The first milestone – 15 months of age

##### 5.1.1 Objective performance criteria

The table presented below details the characteristics of Hal's conversational performance with some examples. A quantitative performance analysis between Hal and two child-parent conversations is presented in the Appendix.

| <i>Characteristics</i>   | <i>Performance examples</i>  |
|--|--|
| Hal experimented with combining two identical monosyllabic utterances into a bisyllabic one                                  | Trainer: ba<br>Hal: <b>ba ba</b><br>Hal: <b>baba</b>   |
| Hal experimented with combining two different monosyllabic utterances into a bisyllabic one                                  | Trainer: mi ba<br>Hal: <b>mi ba</b><br>Hal: <b>miba</b>  |
| Hal's productive vocabulary consisted of 10 words dominated by nouns   | <b>Mommy, ball, block, bunny, birdy, juice, daddy, dolly, hal, and bye bye.</b>  |
| Hal had a passive vocabulary of about 60 words   | Hal relevantly responds to its trainers' input which consisted of at least 60 words  |
| Hal was able to follow the phonetic combinations of its input language and imitate words.                                    | Trainer: say ball<br>Hal: <b>ball</b>  |
| Hal was able to converse about its needs, engage in turn taking games, and name several objects in its immediate environment | Trainer: give me the ball<br>Hal: <b>ball</b><br><br>Trainer: Give bunny the ball<br>Hal: <b>bunny</b><br><br>Hal: <b>Juice</b><br>Trainer: Do you want some juice?<br>Hal: <b>juice</b> |
| Hal's production is functionally driven  | Hal's words relate to its well-being (e.g., juice, milk) and to naming activities (e.g., ball, dolly)  |
| Hal uses both babbling and words in its conversation and had a Mean Length of Utterance of 1.5                               | Hal: <b>all all ball</b>   |

### 5.1.2 Passing the Turing Test for a 15 month old baby converser

An unsuspecting professional psycholinguist reviewed Hal's conversation transcript and concluded that Hal is conversing like a 15 month-old baby. The report included the following impressions: Hal's productive performance included around 10 words, which are typical productions of children at this age. His ability to imitate, and to produce noun-dominated responses are also characteristic. The level of relevancy of his responses and his understanding of requests and questions are age appropriate.

In general, Hal's behavior can be considered normal for a child at this age. This is especially true in light of the fact that some of his performance included typical errors that children make at this age. For example the error described by the expert presented below

Trainer: ok but lets read a story first here's a book say 'book'  
HAL: **block...block...**

#### *Interpretation of the above performance:*

The above error is derived from a phonetic association with the sound /b/. The consonant b and the vowel that follows it in both cases are quite similar for a child, because the /l/ sound between the /b/ and the /o/ is a glide or a semi vowel and it is masqueraded by the /o/.

The child is trying to imitate and the first attempt is a retrieval of the closest phonetic association he can make. Clearly, there are no semantic attributes to 'book' at this stage, which is typical at this age.

## 5.2 The second milestone – 18 months of age

In the second cycle, which is currently ongoing, the system has developed and has become a more mature communicator. A sample conversation corresponding to an 18-month-old human converser (or perhaps even older) is shown below.

### 5.2.1 Sample conversational performance

In their second year of life, children's language performance develops into a two-word combination stage such that words that were used in the one-word stage are now combined into short sentence-like sequences or utterances typically without articles, prepositions, inflections, and other grammatical characteristics required in adult language. Children say things like 'daddy gone' meaning 'daddy is gone' or 'mommy milk' meaning 'mommy, I want some milk'. In the sample below, when Hal says for example 'take car bye bye' he probably means 'let's take the car and go to the park'. A remarkable landmark in Hal's performance is the utterance 'monkeys eat bananas' which Hal remembered from a previous conversation about a visit to the zoo, using it relevantly in the current conversation. The color codes for the sample below are: the black font represents the trainer's input, the blue font represents Hal's output, the red font represents the trainer's modeling, and the green font represents punishments.

## 18-month language development sample

|         |   |
|---------|---|
| Trainer | where is daddy?   |
| Hal     | daddy gone  |
| Trainer | daddy is at work  |
| Hal     | daddy work bye bye  |
| Trainer | yes daddy has gone to work  |
| Hal     | daddy work bye bye bye bye  |
| Trainer | do you want to go to the park?  |
| Hal     | yes bye bye   |
| Trainer | ok we can go to the park  |
| Hal     | take car bye bye  |
| Trainer | we will take the car because the park is far. let's pack some things in a bag |
| Hal     | ball now park   |
| Trainer | yes we should put the ball in the bag   |
| Hal     | ball now park mommy   |
| Trainer | wait hal, we need to take something to eat                                    |
| Hal     | bananastory? night now  |
| Trainer | we need to take something to eat  |
| Hal     | bananas   |
| Trainer | ok we can take some bananas   |
| Hal     | monkeys eat bananas   |
| Trainer | you're so smart hal. do you want to take a book?                              |
| Hal     | juice mommy   |
| Trainer | sure we will take some juice as well  |

## 6 Conclusion

In this paper we have presented the research and development cycle practiced at *Ai*. The motivation of our research is the development of a computer program that can converse in natural language, and it is our firm conviction that a prerequisite for the development of such a system is a commitment to Turing's understanding of intelligence.

We have described the two facets of our approach: a team of researchers who develop general learning algorithms in line with our beliefs that learning, prediction, and compression are inextricably entwined and that behavioristic learning approaches will yield fruit, and a team of trainers who train the developed system to perform, and provide evaluative feedback on the performance relative to equivalent human performance. Objective and subjective assessments of performance are constantly compared to

findings reported in the child language development field. The unique interplay between these two teams has been shown to be successful, resulting in the development of a learning system, which can acquire and use language at the level of an eighteen-month-old human infant. We submit that a developmental approach is a prerequisite to the emergence of intelligent lingual behavior and to the assessment thereof. This approach will help establish standards that are in line with Turing's understanding of intelligence, and will enable evaluation across systems. We predict that the current paradigm shift in understanding the concepts of artificial intelligence and natural language will result in the development of groundbreaking technologies, which will pass the Turing Test within the next ten years.

<sup>1</sup> A.M. Turing, "Computing machinery and intelligence," in *Collected Works of A.M. Turing: Mechanical Intelligence*, D.C. Ince, Ed., chapter 5, pp. 133 (160). Elsevier Science Publishers, 1992.

<sup>2</sup> Stuart M. Shieber, "Lessons from a restricted Turing test," Available at the computation and Language e-print server as cmp-ig/9404002., 1994.

<sup>3</sup> K. Hasida and Y. Den, "A synthetic evaluation of dialogue systems," in *Machine Conversations*, Yorick Wilks, Ed. Kluwer Academic Publishers, 1999.

<sup>4</sup> Noam Chomsky, *Syntactic Structures*, Mouton, 1975.

<sup>5</sup> B.F. Skinner, *Verbal Behavior*, Prentice-Hall, 1957.

<sup>6</sup> R.E. Owens, *Language Development*, Macmillan Publishing Company, 1992.

<sup>7</sup> G. Whitehurst and B. Zimmerman, "Structure and function: A comparison of two views of development of language and cognition," in *The Functions of Language and Cognition*, G. Whitehurst and B. Zimmerman, Eds. Academic Press, 1979.

<sup>8</sup> Claude E. Shannon and Warren Weaver, *The Mathematical theory of Communication*, University of Illinois Press, 1949.

<sup>9</sup> Paul Fletcher & Brian MacWhinney (eds.), "The Handbook of Child Language", 1995, Blackwell Publishers: Cambridge Mass. USA.

<sup>10</sup> <http://childes.psy.cmu.edu/html/clan.html>

<sup>11</sup> A. Goren, G. Tucker, and G.M. Ginsberg, "Language dysfunction in schizophrenia", *European Journal of Disorders of Communication*, vol. 31, no. 2, pp. 467 (482), 1996.

<sup>12</sup> A. Goren, "The language deficit in schizophrenia from a developmental perspective", in *The Israeli Association of Speech and Hearing Clinicians*, 1997.

<sup>13</sup> Goren, A., Fine, J., Manaim, H., & Apter, A. (1995). Verbal and non-verbal expressions of central deficits in schizophrenia. *Journal of Nervous and Mental Disease*, 183, 11, 715-719.