

Intelligent Agent for Acquisition of the Mother Tongue Vocabulary

Bogdan Pătruț

Department of Mathematics and Computer Science,
Faculty of Sciences, "Vasile Alecsandri" University of Bacău
Calea Mărășești, 157, 600115, Bacău, Romania
bogdan@edusoft.ro

Grigor Moldovan

Faculty of Mathematics and Computer Science,
"Babeș-Bolyai" University of Cluj-Napoca
Mihail Kogalniceanu, 1, Cluj-Napoca, 400084, Romania,
moldovan@cs.ubbcluj.ro

Abstract

This paper describes the following: firstly, the basic ideas of a system that simulates how we consider that a child acquires the mother tongue vocabulary and makes the correspondences between objects, words and senses; secondly: the mechanism for a system that can learn the mother tongue vocabulary using observations, and, thirdly, how to make an intelligent agent that can behave like a little child, in the process of mother tongue acquisition.

Keywords: mother tongue, intelligent agent, intelligence

1. Introduction

From his first day a baby learns. He learns by interacting with his environment. He uses his senses to acquire information about the world. Psychologists and doctors claim that not all the five senses have the same importance in the learning process ([4], [5], [6], [7]). They appreciate that the role of sight is more important than other senses. Of course, during the first days of life, a child does not use sight for mental acquisitions. He uses touch, smell and taste. He can recognize his mother by using smell. But the child starts learning the mother tongue when he is two years old. At this age, the role of the senses is changed. Sight is the most important sense, and it is followed by hearing. Throughout our life, 80% of the information is acquired by sight, 18% by hearing and only 2% by the other senses. This is a special capacity of the human beings, who can use articulated language, the role of language being very important in learning and acquiring information.

Human language is intimately related to human intelligence. Humans talk, but no other beings do. The problem is how a baby acquires the mother tongue vocabulary? Our opinion, and not only ours, is that the environment is very important for a human being. We call environment not only the natural environment, but also the family of the child. If parents talk to their child, he will learn the mother tongue. If parents do not talk to their child, he will not learn how to talk.. His vocabulary depends on the parents' vocabulary. For example, we know that children who were born and grew up in forests, don't know any kind of human language. For example, a jungle-child can't speak, and it is very hard to make him speak, if he is over four years old. Tarzan's "mother tongue" is a... jungle language one. [1]

So, the language of a human being is very linked with what he learns from his parents. And parents show and speak to their child. He learns the vocabulary from his parents, and makes correspondences between what they tell him and what they show him, and even what he himself hears and sees in the environment. If you want to learn a foreign language, you need to have good command of your mother tongue.

There is no doubt that babies and toddlers learn a language in a different way than adults do. Until the late 1950's, behavioral researchers were of the opinion that babies are born without any linguistic disposition. As a result of further examinations, it became clear that acquiring a language as well as acquiring human behavior is based on stimulus-response mechanisms. If the child frequently gets in touch with these stimuli and is encouraged, he will learn to repeat the sounds that have triggered a positive reaction in a similar situation. Does this mean, through the eyes of the

behavioral researchers, that second language acquisition concentrates on learning certain language patterns by heart?

One of the most convincing arguments against this theory of children's language acquisition is: If children were able to learn a language only by imitating sounds, they would only be able to use words and sentences that they have heard before. However, babies and toddlers use real words and structures to create new words. As a result of this, researchers came to the conclusion that babies have the ability to acquire language from birth on. The researcher Chomsky has shown that each human being has a disposition to develop language. By seeing how language is used, children finally learn to use it themselves. [2]

2. How a child acquires the mother tongue vocabulary

If a child is alone, he can only see and hear, touch, smell and taste objects and actions or events. But if a child is with his parents, who talk to him, the child will see objects and actions, and he will hear words from the parents.

It is natural for the child to want to learn. So, the child will make associations between what he hears from the parent's mouth and what he sees in the environment. Let's suppose that the mother shows a blue cup to the child, and the mother tells him: *This is a blue cup*. But don't be sure that the child will understand the idea of cup or the idea of blue, from the first example. He can make all kinds of associations between what he sees (a blue cup) and what he hears (the mother's words). He can think that the cup is called cup, but maybe he will think that the cup is called blue. Of course, the cup has a dimension, it may be small or big. Maybe the mother refers to the colour and not to the dimension, but maybe the child observes the dimension and not the colour. We don't know whether the child will associate the word blue or even the word cup with the idea of small or big.

So, the learning process is very complex, and it is a probabilistic one. Thus, a second example is required, and even more than that. Normally, the mother speaks naturally to her child, she speaks with love and affection, she doesn't "judge" or "program" what to say to her child. Therefore, another day she will tell her child, for example, "My darling son, let's *drink* the *milk* from the *cup*."

So, after these two events, the associations can be the following (and not only these) (figure 1):

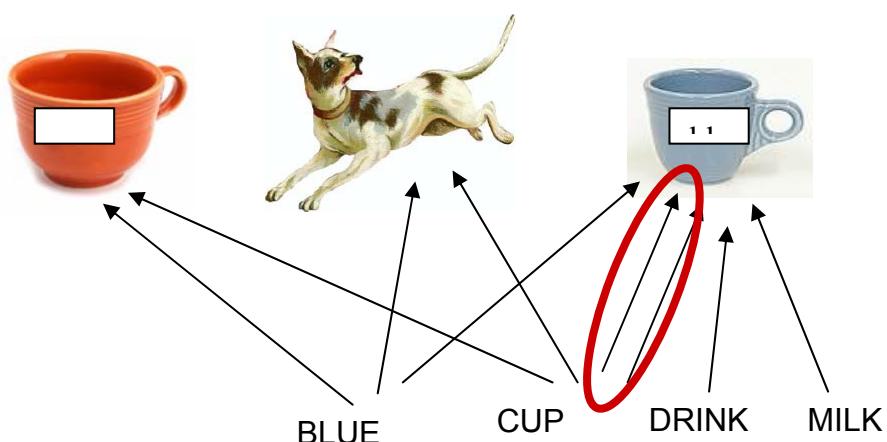


Figure 1. Associations between objects and words

Now we observe that there are two arrows from the word cup and the object cup, and the chance that the idea of cup is represented by the word cup is, probabilistically speaking, greater than other chances. Of course, we cannot draw conclusions only after two examples. But, of course, the correspondence between the concept of cup and the word *cup* will increase, if the word cup occurs many times when the mother shows a cup to her child, even she also shows other objects and she tells him about other things as well, using other words.

The mother speaks to her child every day. We know that the richness of the little child's vocabulary is linked to the richness of the conversations (unidirectional or even bidirectional) between the parents and their child. So, the child will acquire the vocabulary by a probabilistic way. Now the problem is the following: suppose the child has an internal grammar (as Noam Chomsky said, in a discussion with Jean Piaget, at a round table). One may use certain rules, such as the following:

$$\begin{aligned} S &\rightarrow NP\ VP \mid VP \\ NP &\rightarrow N \mid Det\ N \mid etc. \\ VP &\rightarrow V \mid V\ NP \\ &etc. \end{aligned}$$

Does every rule has the same importance, the same "weight" or probability? And, if each rule has its own importance, its own weight, how can we determine this?

3. The mechanism

Let us define a weight grammar through a system $G=(N,T,R,W,S)$, where N is a set of nonterminal symbols, T is a set of terminal symbols, S is a nonterminal symbol from N , R is the set of production rules $N \rightarrow (N^*T)^*$, and W is a function $W:R \rightarrow [0,1]$, which associates a certain weight to every rule. For a rule r from R , we define $|r| = the\ number\ from\ the\ right\ side\ of\ r$. [3]

This weight is, in fact, a probability that says that if we have the rules r_1, r_2, \dots, r_k having the same nonterminal symbol in the left side, and $W(r_1)=w_1, W(r_2)=w_2, \dots, W(r_k)=w_k$, then $w_1+w_2+\dots+w_k \leq 1$, then it will be 1 when the grammar contains all the rules that a natural language may have, with the same nonterminal symbol in the left side.

It is important to determine these weights. We can not make $r_i=1/k$ for each i from 1 to k , we must make correspondances between phrases and observations. We can define a *sample* a couple between a set of objects and actions that the child observes (using sight), and a phrase that his mother tells him, at the same time. A sample, will be a couple (Ob, Ph) , where $Ob=(o_1, o_2, \dots, o_m)$ and $Ph=(u_1, u_2, \dots, u_n)$.

First of all, the child will make a correspondence between what he hears and his internal grammar. Suppose he hears a phrase $Ph=u_1 u_2 \dots u_n$, with n words. He will try to match the phrase with all the rules from R , that have $|r|=|Ph|=n$. Then, for each word u_i he will make the correspondence with all the objects from $O: o_1, \dots, o_m$.

The child can use a matrix A with many rows and columns, and the number of rows will increment each time a new object is seen, and the number of columns will increment each time a new word is heard. In this matrix, $A[o, u]$ represents the number of correspondences between the observed object o and the heard word u . So, at a sample (Ob, Ph) , the child will make this:

```
for each  $o_i$  in  $Ob$ 
for each  $u_j$  in  $Ph$ 
     $A[o_i, u_j]++$ 
```

But, we consider that not only a matrix A is required. For each part of speech, the child must have a matrix. So, suppose he works only with nouns, adjectives and verbs, let A_s be the matrix for nouns, A_a the matrix for adjectives, and A_v the matrix for verbs. The child will increment the matrices A_n, A_a and A_v according to the number of words from Ph . Thus, by using only the rules that have the same cardinal as Ph , where nouns, adjectives and verbs appear, the A matrices will increment their cells, as we have shown above.

We are working on a system that can learn vocabulary through "sight" and "hearing", making correct or almost correct associations between words (nouns, adjectives and verbs) and objects from the environment.

The system can observe different things (objects and actions), and he can make the statistical tables (A_n, A_a and A_v) to find the correct or almost correct correspondences between things and words, using the phrases or just noun phrases that we use as inputs.

The work is currently in progress and we try to finalize it by the end of next year.

The image bellow represents a screen capture from the system. The top-left corner contains the observations (loaded from a file *observ1*). In the top-right corner there is the *An* table, the left-bottom displays the *Aa* table, and in the bottom-right corner there is the *Av* table. Below the observation table you can see the phrases spoken by the parent.



Figure 2. Screen capture of the current application

From this screen capture (figure 2), you can observe that, for example, the noun car (masina) is in a great correspondence with the correct word car(masina), and in the same correspondence with the incorrect word small (mic). But this is due to the fact that the system worked only with few examples. After a training with much more examples, the system will increase the correspondence between the concept/object car and the word car, and the correspondence between the object car and the word little will remain smallest.

If you will look in the second table, where the matrix for adjectives is shown, you can see that the concept of "small" ("mic") has a greatest correspondence with the word "small" ("mic"), and this demonstrates the efficiency of the system. By looking at the top-left corner, you can see that the word "mic" ("small") appears many times. So, the chances for the correspondence between the concept of small and the correct word small ("mic") will grow if the system observes many small objects and the mother uses this word accordingly.

4. Conclusions

Our goal is to make an intelligent agent that can act like a child who acquires the mother tongue vocabulary. It will observe, and it will act. It will act by asking questions in order to obtain much more precise information. For example, he can ask for information that may help him in

making differences between small and car, although one of these words may a noun and the other one an adjective.

We have described, in this paper, the basic idea of how to simulate the way in which a child can acquire his mother tongue vocabulary. We do not know exactly whether this is the way in which a child acts, but our goal is to make an intelligent agent for this kind of acquisition.

Our next task is to improve the system, and to make the system autonomous, so that it may become a real software agent.

5. References

- [1] Kiraly, Don (n.d.). *How do children learn their mother tongue*. Retrieved September 1, 2007, from <http://www.fask.uni-mainz.de/user/kiraly/English/gruppe2/Wie%20lernen%20Kinder.html>.
- [2] Piattelli-Palmarini, M. (1982). *Théories du langage. Théories de l'apprentissage - Le débat entre Jean Piaget et Noam Chomsky organisé et recueilli par Massimo Piattelli-Palmarini*. Paris: Editions du Seuil.
- [3] Moldovan, Gr. (2006). *Limbaje formale și teoria automatelor/ Formal Language and Automata Theory* (Romanian). Bacău, Romania: EduSoft.
- [4] Thielen, Melanie (1996). The Teaching of Short-Term English Courses for German Youths in Great Britain - A Neglected Field of Study: Guidelines for the Teacher. In *Fachbereich Angewandte Sprach- und Kulturwissenschaft der Johannes Gutenberg*, Universität Mainz in Germersheim.
- [5] Malv, H. (n.d.) A World Where Everyone Understands one Another is a Better World. Retrieved January 14, 2008 from <http://www.2-2.se/en/index.html#toc>.
- [6] Komar, S. (n.d.) The Role of the Mother Tongue upon the Acquisition of English Tonality and Tonicity Rules. Retrieved January 14, 2008 from <http://www.phon.ucl.ac.uk/home/johnm/ptlc2001/pdf/komar.pdf>.
- [7] Cummins, J. (n.d.) Bilingual Children's Mother Tongue. Why is it important for education? Retrieved September 14, 2008 from <http://inet.dpb.dpu.dk/infodok/sprogforum/Espr19/CumminsENG.pdf>.