

PLAS, A Paradigmatic Language Acquisition System: An Overview

V. Sembugamoorthy
Computer Centre
Indian Institute of Technology, Powai
Bombay 400 076, India

This paper informally describes the capabilities of a teachable analogy-based language - independent natural language acquisition system known as PLAS. Its language comprehension is intended to resemble that of pre-school children in certain significant aspects. It can be taught, through examples, to understand and acquire the situational aspects (i.e., physical objects, agents, their sensori-motor properties and spatial relations) described in a text. It has no built-in knowledge of English or the domain of discourse. Neither it infers any formal grammar of English.

1. INTRODUCTION

This paper informally describes the capabilities of a primitive natural language acquisition system known as PLAS. The version reported here has been implemented using UCI-LISP on the DEC-10 system at the NCSDCT, TIFR, Bombay. Its linguistic comprehension is intended to resemble that of a pre-school child in certain significant aspects. PLAS can be taught, through examples, to understand and acquire language utterances representing physical objects, agents, their sensori-motor properties (e.g: colour, height, etc.) and spatial relations (e.g: on the chair, to the left of, etc.). To this end, PLAS has to acquire and use the knowledge to analyse and interpret linguistic utterances, understand preforms, obtain the subject of a sentence, and identify and associate the sensori-motor properties and relations with the subject.

PLAS differs significantly from the reported language acquisition systems [3, 4, 5, 6] in one or more of the following aspects: PLAS is neither based on any built-in knowledge of English nor designed to acquire any formal grammar of English. The language understanding and generation procedures are analogy-based. The primitives of the internal representation of language behaviour and world knowledge are either acquired through natural language interaction, or built-in and independent of any language, task or domain.

Within its scope, PLAS provides a solution to the issue of assimilation and accommodation raised in [7].

The machinery of PLAS has been motivated primarily by our long-term goal to model child language behaviour. Nevertheless, the computational machinery should be useful towards constructing practical expandable natural language interaction systems. Refer to [1, 2] for the psychological aspects of PLAS, the explicit substantiation of its features mentioned above and the detailed discussion of its computational machinery.

2. AN EXAMPLE OF INTERACTION

Initially PLAS does not possess any knowledge of English or the domain of discourse. Using a simulated teaching environment called 'Teach-names', the teacher first associates a set of names with the physical objects and agents that he proposes to deal with in his interaction. Note that neither all names of an object/agent nor all objects that a word or a phrase can name need be associated using 'Teach-names'. A typical input to 'Teach-names' is:

1. CHAIR 1, CHAIR2; CHAIR; (OBJECT MADE-OF-WOOD).

This means that the language expression 'CHAIR' is a name of the physical objects 'CHAIR1' and 'CHAIR2'; '(OBJECT MADE-OF-WOOD)' is called a taglist. It is intended to simulate the structured complex of sensori-motor mechanisms which is responsible for the perception/manipulation of 'CHAIR1' and 'CHAIR2'. However, since articulating this structure is itself a challenging open problem, a taglist is presently assumed to be an unstructured set of tags where each tag relates to a sensori-motor mechanism. As we shall see in section 3, the

taglist plays an important role in understanding linguistic utterances. Moreover, it is significant to note that the language expressions used to name tags (e.g: MADE-OF-WOOD) do not function as language expressions as far as PIAS is concerned. Let us assume that using 'Teach-names', the teacher has taught PIAS the following words and phrases:

2. TABLE, THE TABLE; 3. WINDOW; 4. DESK;
5. ROOM, THE ROOM; 6. FOUNTAIN, THE FOUNTAIN; 7. BOX; 8. SHELF.

Next, the teacher teaches, using another teaching device called 'Teach-aspects', a set of words and phrases known as aspect tokens as shown below (Aspect tokens are language expressions that denote properties of, and relations between, objects and/or agents):

9. NAC (BROWN; (COLOUR)).
This means that 'BROWN' is an aspect token of a New Aspect Category (NAC for short). 'COLOUR' is the taglist associated with 'BROWN' at the sensori-motor level.
10. (GREEN; (COLOUR)) SAC (BROWN; (COLOUR)).
The aspect tokens 'GREEN' and 'BROWN' belong to the Same Aspect Category (SAC).

11. (RED; (COLOUR)) SAC (GREEN; (COLOUR)).
12. (RED IN COLOUR; (COLOUR)) SAS (RED; (COLOUR)).
'RED IN COLOUR' and 'RED' represent the Same Aspect (SAS). Let us assume that using 'Teach-aspects', the teacher has taught the following aspect tokens:

13. PLASTIC; 14. WOOD; 15. NEAR THE FOUNTAIN; 16. IN THE MIDDLE OF THE FOUNTAIN; 17. INSIDE THE DESK; 18. ON THE BOX; 19. MADE OF PLASTIC; 20. HEAVY
21. LIGHT.

Next, the teacher uses a teaching device called 'Teach-Proforms' to teach proform tokens as shown below:

22. (THE CHAIR NEAR THE TABLE) ASS (THE ONE NEAR THE TABLE).
This means that 'THE CHAIR NEAR THE TABLE' is an ASSignment (ASS for short) of the proform 'THE ONE NEAR THE TABLE'.

23. (THE CHAIR) ASS (IT).
Before we proceed further, it is worthwhile to point out that (i) in principle, 'Teach-names', 'Teach-aspects' and 'Teach-proforms' can be employed in any order and any number of times; (ii) the teaching devices 'Teach-aspects' and

'Teach-proforms' are unrealistic from the view-point of modelling the learning environment of pre-school children and can be eliminated, as shown in [1], by incorporating, in PIAS, a question-answering environment.

Next, the teacher teaches a text (call Text1) as shown below:

24. ((A CHAIR) REF 'CHAIR1' AND (A TABLE) REF 'TABLE1') REF (COLLECTION1 AGG CHAIR1, TABLE1) ARE NEAR THE WINDOW.

This means that the tokens 'A CHAIR', 'A TABLE' and 'A TABLE AND A CHAIR' refer to 'CHAIR1', 'TABLE1' and 'COLLECTION1' respectively; 'COLLECTION1' consists of 'CHAIR1' and 'TABLE1'. The sentence taught is 'A CHAIR AND A TABLE ARE NEAR THE WINDOW'. REF and AGG, which stand for 'reference-of' and 'aggregate-of', simulate the extra-linguistic means employed by the language community to associate a language expression with its referent, and to specify the members of an aggregate of objects/agents respectively. Besides interpreting these extra-linguistic markers, PIAS has to analyse and interpret 'NEAR THE WINDOW' using 15 i.e., step 15; identify 'A CHAIR AND A TABLE' as the subject token; associate the aspect token 'NEAR THE WINDOW' with 'COLLECTION1' and thereby with 'CHAIR1' and 'TABLE1'.

25. THE CHAIR IS BROWN IN COLOUR.
Using 1 and 2 i.e., using the knowledge acquired in 1 and 2, PIAS interprets 'THE CHAIR' as 'CHAIR1'; acquires 'BROWN IN COLOUR' using 12; identifies the subject token 'THE CHAIR'; associates 'BROWN IN COLOUR' with 'CHAIR1'.

26. THE TABLE IS RED IN COLOUR.
PIAS uses 25 to understand this sentence since it is similar to the one in 25.

27. A DESK AND A CHAIR ARE IN THE MIDDLE OF THE ROOM.
'A CHAIR' is interpreted as 'CHAIR2' because 'CHAIR1' has been associated with 'NEAR THE WINDOW' in 24. 'IN THE MIDDLE OF THE ROOM' is understood using 16.

28. THE BROWN CHAIR IS MADE OF WOOD.
PIAS interprets 'THE BROWN CHAIR' as 'CHAIR1' and not as 'CHAIR2'. Note that 'CHAIR1' has been associated in 25 with 'BROWN IN COLOUR'. PIAS uses 12 to equate 'BROWN IN COLOUR' and 'BROWN', and to correctly interpret 'THE BROWN CHAIR' as 'CHAIR1'.

29. THE ONE IN THE MIDDLE OF THE ROOM IS MADE OF PLASTIC.
PIAS assigns 'THE CHAIR IN THE MIDDLE OF THE

ROOM' to the proform 'THE ONE IN THE MIDDLE OF THE ROOM'. Note that the teacher has taught, in 22, only the assignment 'THE CHAIR NEAR THE TABLE'. Moreover, PLAS has to use 27 to interpret 'THE CHAIR IN THE MIDDLE OF THE ROOM' as 'CHAIR2' and not as 'CHAIR1'.

The teacher next tests PLAS by asking it to understand a similar new text. For example,
Text 2

THIS IS A CHAIR. THERE IS A BOX ON THE CHAIR.
IT IS HEAVY. THE BOX INSIDE THE SHELF IS
LIGHT. THE HEAVY BOX IS GREEN IN COLOUR.

Note that PLAS can understand Text2 even though the underlying sentential forms and the described situational aspects vary significantly from those of Text1. Moreover, PLAS should interpret 'IT' in the third sentence as the box referred to in the second sentence even though the teacher has associated only 'THE CHAIR' with 'IT' in 23. Now, as requested by the teacher, PLAS generates the following sentences using 'CHAIR' and 'RED':

THE HEAVY CHAIR IS RED IN COLOUR. THE LIGHT
CHAIR IS RED IN COLOUR. THE RED CHAIR IS
MADE OF WOOD. THE RED CHAIR IS MADE OF
PLASTIC. THE CHAIR IS RED IN COLOUR.

Note that none of these sentences appears in Text1 or Text2. Moreover, PLAS has not generated any semantically anomalous sentence like 'THE RED CHAIR IS GREEN IN COLOUR'.

3. THE COMPUTATIONAL MACHINERY

Interaction procedures simulate the teaching and testing environments discussed in section 2. They invoke sensori-motor procedures and language behaviour procedures to process the sensori-motor level information (i.e., physical objects, taglists, extra-linguistic markers) and the linguistic expressions respectively. In what follows we shall sketch only the language understanding procedure. To understand a language expression, say 'ON THE TABLE', PLAS first invokes the segmentation procedure to analyse it into a known language schema, say 'on -' (' - ' stands for a variable) and the filler 'THE TABLE'; 'THE TABLE' itself has to be understood before it is accepted as a filler. The segmented language expression is then validated. For instance, if 'NEAR THE CHAIR' is an already known valid filler for the second variable of the schema '- IS -', then 'ON THE TABLE' is accepted as a valid filler for the same variable provided the taglists associated with the two aspect tokens match (presently two taglists match iff they contain identical tags). The interpretation procedure then interprets the validated token. For example, using the knowledge that the referent of 'JOHN'S BOOK' is

say 'BOOK10', 'JILL'S TABLE' will be interpreted as, say 'TABLE5'. A successfully segmented, validated and interpreted language expression is said to be understood. If the schema 'On -' is not known already, it will be abstracted out of it by the schematization procedure by replacing understandable or already known language expressions in it. In this case also, before 'ON THE TABLE' is declared understood, it has to be validated and interpreted. Note that what underlies the language understanding procedure is not any language-specific information but are the capabilities to assimilate the unknown in terms of the known, and to accommodate the unknown if assimilation fails. These capabilities are analogy-based. To enable these capabilities is the main function of the data spaces of PLAS.

ACKNOWLEDGEMENTS

The valuable guidance of Prof.R.Narasimhan is gratefully acknowledged.

REFERENCES

- [1] Sembugamoorthy, V. 'A Paradigmatic Language Acquisition System', Ph.D.Thesis, Indian Institute of Tech., Bombay, India, May, 1978.
- [2] Narasimhan, R. 'Modelling Language Behaviour', Tech.Report, Computer group, TIFR, April, 1973.
- [3] Jordan, R.S. 'A Natural Language Understanding Based on a Freely Associated Learned Memory Net', Int.J.Comput.Inf.Sci.6:1 (1977) 10-25.
- [4] Reeker, L.H. 'The Computational Study of Language Acquisition' in Advances in Computers, Rubinoff, M., and Yovits, M.C., (Eds) Academic Press, 1976, pp.181-235.
- [5] Siklossy, L. 'Natural Language Learning by Computer' in Representation and Meaning: Experiments with Information Processing Systems, Prentice-Hall, NJ, 1972, pp.288-328.
- [6] Harris, L.R. 'A System for Primitive Natural Language Acquisition', Int. J.Man-Mach. Stud.9 (1977) 153-206.
- [7] Moore, J. and Newell, A. 'How can Merlin Understand?' in Knowledge and Cognition, Gregg, L. (Ed), Lawrence Erlbaum Assoc. Potomoc, Md., 1973.