Personal Details

Principal Investigator	Prof. A. Raghuramaraju	Department of Philosophy, University of Hyderabad
Paper Coordinator	Dr. Geeta Ramana	Department of Philosophy, University of Mumbai
Content Writer	Dr. Smita Sirker	Department of Philosophy Jadavpur University
Content Reviewer	Prof. Nizar Ahmad	SSUS, Kalady
Language Editor	Miss. Chitralekha D R	Freelancer, Chennai

Description of Module

Subject Name	Philosophy
Paper Name	Philosophy of Language
Module Name / Title	Language of Thought and Language of Artificial Intelligence
Module Id	11.22
Pre-requisites	Preliminary understanding of mental representation
Objectives	To provide a very basic introduction of Language of Thought and Language of Artificial Intelligence
Key words	Mental representation, artificial intelligence, language of thought, symbol manipulating system, computational representational theory of mind, linguistic representation

Language of Thought and Language of Artificial Intelligence

1. What is Language of Thought?

Kim Sterelny observes "The language of thought hypothesis is an idea, or family of ideas, about the way we represent our world, and hence an idea about how our behaviour is to be explained" (Sterelny 1999, 451). We, who are newly introduced to this idea or hypothesis, would find it fascinating when we try to decipher the apparent meaning of the phrase "language of thought", for it may appear that it hints at some form of language which is the language of our thought or thinking. What is meant by such a language and how do we have it? Let us begin with the above observation. Human beings are said to be highly adaptive to their surroundings, making them highly efficient to find their ways within the maze of social and physical environments and challenges. Each day of our lives is a mesh of decisions ranging from simple to highly complex tasks, done almost on a regular basis. Our ability to negotiate and adapt to our complex environment(s) is due to our mental capacities. Sterelny further writes "We navigate our way through our social and physical world by constructing an inner representation, an inner map of that world, and we plot our course from that inner map and from our representation of where we want to get to. Our capacity for negotiating our complex and variable environment is based on a representation of the world as we take it to be, and a representation of the world as we would like it to be"(Ibid.). In other words, we have an internal representation of the world in our minds. For instance, I have a mental map of the roads that I must take in order to commute from my residence to the place where I work. Based on this internal representation of the physical routes between the two destinations, I choose to follow a particular road map on my way to office. On a day, where the regular road is closed, I can decide on an alternate route provided I have a prior representation of the alternate route. Otherwise, I depend on the advice of someone else.

Here what is important to note is that given the idea that we have an internal representation of the world in our minds and our beliefs, desires and actions are shaped by ways in which the world is represented in us; there have been debates regarding the nature of these mental representations. The language of thought (henceforth, LOT) hypothesis is the hypothesis that our mental representation has a linguistic structure. Our thinking and thoughts are said to take place in a mental language. Schneider explains, "According to the LOT program, conceptual thinking occurs in an internal language-like representational medium. However, this internal language is not equivalent to one's spoken language(s). Instead, LOT is the format in which the mind represents concepts. The LOT hypothesis holds that the mind has numerous mental 'words' (called *symbols*) that combine into mental sentences according to the grammatical principles of the language. When one thinks, one is engaged in the algorithmic processing of strings of these mental symbols" (Schneider 2011, 6–7).

The LOT hypothesis was first proposed by Jerry Fodor, an American philosopher and cognitive scientist, in his book *The Language of Thought* (1975). According to Fodor, human cognition involves mental representations which are structured like sentences in a language. He proposed that thought and thinking is carried out by a mental language, called *mentalese*, which is different from our natural languages, such as Hindi, English, or French. LOT is the inner mental language that is responsible for the acquisition of language and concept learning. Our mental language or *mentalese* consists of a compositional symbolic or representational system with semantic content, governed in their composition by syntactic specifications. This gives human cognition a linguistic structure.

Fodor's hypothesis became a matter of serious research both in philosophy of mind and philosophy of language. In the next two sections, we briefly sketch why LOT hypothesis is connected to both researches in language and mind. Schneider observes that the LOT program offers "an influential theory of the nature of thought and the minds that have them. With respect to minds, the program says that they are symbol-manipulating devices of an ultra-sophisticated sort. With respect to mental states, these are said to be mental symbols – ways in which we conceive of the world – strung together by an inner grammar, the behaviour of which is to be detailed by a completed cognitive science"(Schneider 2011, 1).

1.1. Connection with Philosophy of Mind

One of the most influential theories of mind in both philosophy of mind and cognitive science is the computational representational theory of mind (henceforth, CRTM). The emergence of this theory was mainly boosted by Fodor's famous *Language of Thought*. According to CRTM, mind is understood as a computational device. The mind functions as a computer and mental functions are nothing over and above computational functions encoded in a language of thought. Thus, the mind is nothing but a formal system of mental representations on which multiple computations can be executed. The representational theory of mind (henceforth, RTM) proposes that the mental states such as beliefs, desires, and so on, are mental representations of the states of affairs or objects they are about. In other words, if you believe Y, then the proposition Y that you believe (the object of your belief) is a representation of something in the world. If you believe that "Today is Monday", then the proposition (of your belief) is a representation of the fact that today is Monday. In simpler terms, our mental states are representations of the world. Suppose,

(a) Sunil believes that tulsi is a beneficial herb. (Belief T = tulsi is a beneficial herb)

Here Sunil's belief T that tulsi is a beneficial herb *represents* (R) the fact (F)[1] that tulsi is a beneficial herb. Thus, T is a representation (R) of the fact (F). In other words, the belief (T) has a representational relation (R) with the fact (F). Now what does it mean to say that T *represents* F? The belief (a mental state) T is a mental representation (R) of the fact (F) means that the mind (which has the belief state T) provides a map of the fact (F) in the world. This mental map is a direct representation of that particular state of affair in the world. Thus, mental representation is some kind of a mental map of the world, as the mind navigates through the world through its inner mechanism consisting of language of thought, which is the inner mental language that expresses the mental representations.

In "Computing Machinery and Intelligence" (Turing 1950, 433–60), Turing proposed that symbol-processing devices can think. According to Fodor, his theory of LOT was inspired by Alan Turing's idea that computation can be defined in terms of formal manipulation of uninterpreted symbols by developing appropriate algorithms. For him, the basic symbol structures in the mind that carry out information processing are sentences in an internal language of thought and information processing works by transforming those sentences in the language of thought (Bermúdez 2010, 156). Thus, LOT manifests a particular philosophical view about mind and its nature of thoughts, that is, mental thoughts possesses as linguistic like structure. Thinking takes place in a mental language, in which symbolic representations are manipulated in accordance to the rules of a combinatorial syntax.

1.2. Connection with Philosophy of Language

Philosophers who advocate the LOT hypothesis posit that "our powers of mental representation to be strikingly similar to our powers of *linguistic representation*" (Sterelny 1999, 452). The richness of our cognitive ability is comparable to our linguistic ability. Sterelny states that both language and thought are not stimulus bound as we can speak or think "of the elsewhere and the else when" (Ibid.). Furthermore, both language and thought are counter-factual as we can speak and think about the world in ways which are not the case. For example, we can think and talk about angels.

The power of linguistic representation is said to be in the organisation of language. Sterelny explains,

> Sentences are structures built out of basic units, words or morphemes. The meaning of the sentence - what it represents - depends on the meaning of those words together with its structure.

[1]It is important to note that mental representations may be of either actual state of affair (a fact) or a possible one. Consider this: "If you believe (Y) that the earth is flat, then your belief (Y) is a representation of the state of affair that the earth is flat." The distinction is primarily between a possible but not existing state of affair and an existing state of affair.

> So when we learn a language, we learn the words together with recipes for building sentences out of them. We thus acquire a representational system of great power and flexibility, for indefinitely many complex representations can be constructed out of its basic elements. Since mental representation exhibits these same properties, we might infer that it is organized in the same way. ... A minimal language of thought hypothesis is the idea that our capacities to think depend on a representational system, in which complex representations are built from a stock of basic elements; the meaning of complex representations depend on their structure and the representational properties of those basic elements; and the basic elements reappear with the same meaning in many structures. This representational system is "Mentalese." (Ibid.)

Sterelny makes it clear why LOT is also an area of interest for philosophy of language. Mental representation is proposed to have a structure akin to that of language, possessing a linguistic structure. A representational system can have a linguistic structure if it employs a combinatorial syntax and a compositional semantics and the LOT hypothesis drives the idea that mental representation has both.

2. Basic claims of Language of Thought (LOT) Hypothesis

In order to apprehend the thesis that thinking is an algorithmic manipulation of mental symbols, let us understand some of the main claims of LOT hypothesis.

(i) LOT claims that mental representation has a linguistic structure. For a representational system to have a linguistic structure, the former requires both combinatorial syntax and compositional semantics (Katz 2015). Therefore, it is posited that mental representation has both and thereby thoughts are said to occur in a formal mental language, i.e. *thoughts are literally sentences in the head*. Hence the system of mental representation is called *language of thought*. Combinatorial syntax and compositional semantics could be found in formal languages. Let us consider propositional logic. Propositional logic employs symbols like A, B, C, etc. to represent simple declarative sentences and symbols like '·', 'v', '→' for logical connectives and, or and if…then, respectively. For instance, A may stand for the atomic representation of the sentence "John is in London', B for an atomic representation of the sentence "Luke is in Norwich" and C for an atomic representation of the sentence "Alice is in Cambridge". Now, [(A · B) v C] will be a compound representation of the sentence 'Either John is in London and Luke is in Norwich, or Alice is in Cambridge'. This compound representation constitutes of a compound representation '(A · B)' and an atomic representation 'C'. Propositional logic uses both atomic and compound representations and the constituents of a compound representation can be either atomic or compound. Thus, propositional logic has a *combinatorial syntax*. Moreover in propositional logic, the semantic

content of such a representation is a function of the content of the syntactic components along with the structure and arrangement of the sentential representations. In other words, compositional semantics is a semantics which sees the semantic value of an expression as composed out of the semantic values of the component-expressions of the expression.

However we must also remember that, LOT posits that this language is not equivalent to any particular language, though it possesses a common linguistic structure in all human thoughts. Fodor claims that our learning of any language requires an internal mental language commonly possessed by us and this mental language is not introspectively accessible to us.

The LOT has a structure composed of symbols with innate rules encoded in us. As a mental language it is structured to express our thoughts. It has a logically articulated syntactic structure which is combinatorial in character, composed of simple symbols. It also has a compositional semantic structure. In other words, the meaning of the complex mental sentences is dependent and derived from the meaning of its component (constitutive) parts. As mentioned earlier, it was Fodor's LOT that led to the emergence of CRTM. For CRTM, thoughts are complex symbols which have both syntactic and semantic properties. Thus, the thought "Ted likes mangoes" is a complex symbol made up of basic symbols such as "Ted", "likes" and "mangoes".

(ii) Systematicity is one of the basic properties of language and it entails that sentences have structures, i.e. systematicity is preserved only if the sentences in a language have a structure. For instance, if it is meaningful to say "X forgives Y", then it is also meaningful to say "Y forgives X". LOT being a symbol system is also characterised by systematicity in the sense that there are features which make sentences logically connected with other sentences. Consider the mental tokens "Bob forgives Ted" and "Ted forgives Bob". These tokens are logically connected since if one of them is meaningful then the other is also meaningful.

(iii) Productivity is another feature of language. Languages are said to be productive in the sense that we can build new meaningful sentences out of parts of sentences or old sentences. For instance, if you have two sentences, "Bhuban lives in Mumbai" and "Bhuban has a Ford" then you can obtain a new sentence "Bhuban lives in Mumbai and has a Ford". Productivity of language is also due to the fact that languages are structured. LOT also has the feature of productivity since this language can also generate complex sentence structures out of simple sentence structures ad infinitum. Two simple mental tokens can generate a further new complex mental token. For instance, if you believe that "Today is Friday" and "Today is pay-day" then a new mental token can be generated – "Today is Friday and pay-day". Thoughts

themselves are of a productive nature since you can build or generate new thoughts from a given set of thoughts.

3. Arguments in favour of LOT Hypothesis

In this section, we will discuss only four important arguments forwarded in favour of the LOT hypothesis. Advocates of LOT argue to justify why we need a language of thought and evidences for LOT are mainly both psychological and linguistic. Thus arguments put forward in favour of LOT concern both language and mental representations. LOT [the set of *mentalese* along with some structural rules for building a representation of the outer world] has close similarities with what are ordinarily called language in respect of being a language. Hence some of the following arguments draw parallel from the features of natural language.

First, as seen before, systematicity is preserved only in a well-structured language. A representational system possesses the property of systematicity when the ability of the system to express certain propositions is intrinsically related to the ability the system has to express certain other propositions (Katz 2015). Now, how does this property act as an argument in favour of LOT? To put it in simple terms, language and thought are said to be systematic. There is a systematic relationship that holds between sentences/thoughts. Combinatorial syntax allows the form of a sentence to be distinct from its meaning. So, in a language, you can make a meaningful sentence with a certain form of syntax and also you can use the same syntactic form to create another new sentence with a different meaning. Like, if you create a sentence “Bob forgives Ted” then you can use the same syntactic form to create another sentence with a different meaning, “Ted forgives Bob”. Fodor and Pylyshyn (1988) argue that if thought is largely systematic then it must be linguistically structured. So if you entertain a thought “Bob forgives Ted”, then you can also entertain the thought “Ted forgives Bob”.

Second, as also discussed earlier, productivity is a feature that is possible in a language that is well-structured. In other words, it can also be said that in principle, a system of representation possesses the property of productivity if it can produce an infinite number of distinct representations. A productive representation system with a finite number of atomic representations can generate an infinite number of compound representations. Again the question arises, how does this property act as an argument in favour of LOT? The argument rests on the assumption that language and thought are productive, that is, an infinite number of sentences/ thoughts can be produced (generated) using combinatorial syntax. Fodor and Pylyshyn (1988) argue that our mental representation is productive. This is because our LOT is embedded in a system possessing combinatorial syntax and compositional semantics. They claim that natural languages are productive. For example, in English there is only a finite number of words but since there is no upper limit on the length of a sentence, there is also no upper bound on the number of unique

sentences that can be generated. In principle, a competent English speaker has the capacity to produce an infinite number of unique sentences. Consider this in the realm of thought. Humans can entertain an infinite number of thoughts generated on the basis of atomic mental tokens (thoughts); that is, compound thoughts can be generated from the atomic ones. Katz summarises that, "Thus, they (human beings) must possess a system that allows for construction of an infinite number of thoughts given only finite atomic parts. The only systems that can do that are systems that possess combinatorial syntax and compositional semantics. Thus, the system of mental representation must possess those features" (Katz 2015).

Third, when we learn a natural language, we need a prior knowledge of another natural language, which we call the first language. Going by principle, we need to learn a language prior to our learning of our first language and so on ad infinitum. This would initiate an infinite regress. In order to terminate this regress, it is argued that we must be endowed with a language which is innate and not learned. This innate language is the LOT. This is the reason why LOT is not a natural language. LOT is called *mentalese* and is innate and universal. This *mentalese* form the basis of all our language learning.

Lastly, when we learn a concept, say "bird" we also expect to have a definition for that concept, so that we understand what objects do fall within that concept class. In order to have this kind of understanding, we need to generalise the form "a is a chair *iff* a has B". Thus concept learning in a way also involves our mastery of generalisation. According to Fodor, human beings have the conceptual capacity to understand chairs, birds, trees, and so on. We do not learn new concepts; instead we learn to put familiar innate concepts together into new combinations. Our learning a natural language is learning to identify the words of the natural language with our innate concepts. We cannot acquire the word for a concept without already having the concept in mentalese.

4. Arguments against the LOT Hypothesis

Though the LOT hypothesis garnered lots of attention, it is not without criticisms. Out of the many objections against the LOT hypothesis, we will discuss two such arguments.

First, proponents of the LOT hypothesis claimed mentalese as an innate language and argued that it was a prerequisite for the learning of any natural language. Thus, LOT could explain how natural languages are learned, how they are understood by us and how the utterances in such languages can be meaningful. For instance, Fodor posits that natural languages are learned by forming and confirming hypotheses about the translation of sentences in natural language (say, English) into mentalese. Thus, a sentence "The colour of the ball is yellow" is true in English *iff* S, where S is a sentence in one's LOT. This translation requires a representational medium to form and confirm hypotheses (to represent the truth-conditions of natural language sentences). This representational medium is LOT.

The argument against such a position is that this generates a regress. Just as in the case of learning and understanding natural languages, one also needs to explain how LOT is learned, how it is understood, and also, how sentences in LOT can be meaningful. This will eventually lead to a regress as the dependence of learning of one language continues to lean on another. Furthermore, if we get a successful explanation for LOT that does not lead to a regress, then it could and ought to be given for any natural language without introducing a LOT.

Fodor responds by arguing how it is different from any natural language. LOT is not learned, it is innate. Furthermore, LOT is understood in a different sense than what is involved in the comprehension of a natural language, and sentences of LOT do not derive their meaning in relation to meaningful sentences in some other language, but in a completely different way involving some sort of a causal relation to what they represent.

Second, Dennett puts forward the following example to argue against Fodor's thesis, "In a recent conversation with the designer of a chess-playing program I heard the following criticism of a rival program: 'it thinks it should get its queen out early.' This ascribes a propositional attitude to the program in a very useful and predictive way, for as the designer went on to say, one can usefully count on chasing that queen around the board. But for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with "I should get my queen out early" explicitly tokened. The level of analysis to which the designer's remark belongs describes features of the program that are, in an entirely innocent way, emergent properties of the computational processes that have "engineering reality." I see no reason to believe that the relation between belief-talk and psychological talk will be any more direct" (Dennett 1981, 107). Dennett tries to argue that it is possible to have propositional attitudes without explicit representations.

The designer programmer ascribes a propositional attitude (the rival program thinks that it should move its queen out early) to a rival program, where this ascription is both useful and predictive. For instance, when we want to program the chess computer to play with its rival computer (program), then we would want the chess computer to produce a defence that is adequate to this ascription. However, if we look at the nature of how chess programs (computer) work, then we know that within the program code of the rival program, there is actually no internal representation of the propositional attitude "should get its queen out early".

Fodor responds that the objection is not that the program has a dispositional belief "should get its queen out early". The program actually operates on this belief, just like when we reason we often follow rules of inference such as modus ponens, hypothetical syllogism, and so on without explicitly representing them. We need to draw a distinction between "rules on the basis of which mentalese data-

structures are manipulated" and "the data-structures". According to Fodor, data-structures have to be explicitly represented when they are formally manipulated by rules. But, we need not have all the rules being explicitly represented in mentalese. Some rules are hard-wired to the system and thereby implicit; some may not be so.

Thus, to conclude, LOTH is a bold hypothesis originally proposed by Fodor which posits that the medium of thought is an innate language that is distinct from the other spoken languages. This innate language, called *mentalese* contains all necessary conceptual resources required for any of the propositions that we can think, understand or express.

5. Language of Artificial Intelligence

The field known as Artificial Intelligence (henceforth, AI) has a long history and is still constantly and actively developing. AI is primarily concerned with the development of computational methods for accomplishing aspects of human intelligent behaviour. AI also aims to arrive at a general theory of intelligent action in agents, which does not only include humans and animals. AI rests on the view that cognition is computational and that the mind and brain are computers. AI researchers believe that computation can be developed that simulates and even explains the working of the human mind. In order to develop such models, AI research requires programming languages that will assist in developing and programming such computational models. AI has been creating many such special languages.

AI systems are constructed to perform tasks similar to what human beings perform through some mental activity. However, the AI systems are nothing but (electronic) computers/machines designed to perform specific type of tasks. But in contrast to the human case, in AI systems, the tasks are performed only mechanically (without use of any human intelligence), algorithmically (in technical terms, effectively) and in finite steps – generally put in modern terms, computationally. The task has to be presented to the system through a language understandable to the (electronic) computer (machine) designed for the task. When presented with a task, the machine "computes" the task through manipulation of symbols – through which the task was presented in the first place – under clear instructions designed for arriving at the result, in a finite number of steps. So, the machine/computer needs to be equipped with a language to perform the task, that is, to receive the task, to proceed towards the completion of the task through manipulations of the symbols with which it is equipped and in which he task has been presented to it. Such languages are the languages for AI systems. These are languages in many respects, no doubt, in sharing some common features with our ordinary languages, but these are very unlike the ordinary languages in many other respects, because these are to be used by electronic machines which work, ultimately, only in two modes – electrons passing through a circuit and not passing through a circuit.

Languages designed for giving a task to the machine are called programming languages. These are what are called languages of AI systems.

Computer programs, such as PROLOG and LISP are used in AI research. One aspect of such AI programming languages is that they provides the ability to implement a physical symbol system. In 1976, Allen Newell and Herbert Simon proposed the physical symbol system hypothesis. They proposed a set of properties that characterise the kind of computations that the mind depends on. According to this hypothesis, intelligent actions must rely on nothing more that syntactic manipulation of formal symbols. Thus, for them, a physical symbol system is both necessary and a sufficient condition for intelligent actions. This hypothesis tried to address the issue regarding the kind of operations that are required for intelligent actions. However, this hypothesis could only be empirically proved or disproved. AI research has been testing this physical symbol system hypothesis. Glasgow and Browse observes that, "This basic ability to retain and transform symbolic structures is generally viewed as central to the development of any intelligent system. While most programming languages center on an ability to manipulate numeric data, AI systems typically exploit knowledge of concepts through their representation as symbolic structures. […] An ideal AI programming language should provide mechanisms for the expression and manipulation of real world knowledge. This is normally done using a logic formalism that allows inferences. To be effective such an AI language must contain a standardized control mechanism and, at the same time, be conducive to the development of improved control and inference methods" (Glasgow and Browse, 1985, 431).

5.1 Philosophical Problems about Language of AI:

Since, language for AI is a language, although designed for a very special purpose, it will lend itself to certain general philosophical questions that are common to any language. For instance, the question about the categories that the expressions of such a language can be divided into, and the consequent question about what type of entities do the expressions falling into different categories refer to will be very important. In short, a projected semantics of the language will be called for. In this case, however, the answers may be somewhat unfamiliar.

The language of AI is supposed to instruct the computer to do certain computing on a given set of information/data. Hence:

(a) Typically, the given set of information/data, in the context of AI, will be nothing but sequences of symbols recognizable by the computer. Such sequences of symbols are supposed to stand for some or other indicative sentence of a part of an ordinary language capable of expressing states of affairs of a domain of discourse for which an AI is being attempted to be

built. Language of AI must have well defined categories of expressions referring to the different types of sequences of symbols the computer recognizes.

(b) Language of AI must also have certain rules of transformation applicable on the sequences of symbols being referred to, to facilitate the required computation on the given data. The rules of transformation, in this case, can only be rules for rewriting sequences of symbols into sequences of symbols of the alphabet of the same language. The final rewriting for a given computational task gives us the result.

(c) The rewriting rules will be algorithmic in the sense that a flowchart can be given for the task that the given AI is built for.

Given the above, language of AI can be seen to be a formal algorithmic system, i.e., formal language with a deductive apparatus attached to it, along with a set of instructions, formally recognizable, on when to apply which rule for deduction. Such a language will be close to what can be found in formal axiomatic systems for logics, but different in that, that this system may deal with some domain of discourse different from that of logic.

Now, the philosophical questions specific to such a language of AI may be the following. An AI is thought to be an artificial automated simulation of human thought processes for various tasks. The question that becomes important is whether the assumptions about so-called mind and mental processes that are used within the construction of such a language of AI, indeed, reflect what goes on in the human performances of the tasks. The various proposed models of how the mind works are to be carefully examined to see whether they come close to the mechanisms of a language of AI. Moreover, it is important to enquire whether different AI systems, along with their associated language of AI, for simulating different human performances are essentially assuming different models for mind and its processes. If the answer is yes, then it will be philosophically an uphill task and almost an extremely difficult challenge to face.