DOCUMENT RESUME

ED 115 081 FL 007 167

AUTHOR Wilks, Yorick

TITLE Philosophy of Language. Course Notes for a Tutorial on Computational Semantics.

PUB DATE Mar 75

NOTE 51p.; Course given at the Institute for Semantic and Cognitive Studies, Switzerland, March 17-22, 1975; For related document, see FL 007 164

AVAILABLE FROM Institute for Semantic and Cognitive Studies, Villa Heleneum, 6976 Castagnola, Switzerland ($10.00)

EDRS PRICE MF-$0.76 HC-$3.32 Plus Postage

DESCRIPTORS Applied Linguistics; *Artificial Intelligence; *Computational Linguistics; Linguistic Theory; *Logic; Logical Thinking; Mathematical Logic; *Philosophy; *Semantics; Syntax; Thought Processes; Transformation Generative Grammar

IDENTIFIERS *Computational Semantics; Montague (Richard); Wittgenstein (Ludwig)

ABSTRACT

This course was part of a tutorial focusing on the state of computational semantics, i.e., the state of work on natural language within the artificial intelligence (AI) paradigm. The discussion in the course centered on the philosophers Richard Montague and Ludwig Wittgenstein. The course was divided into three sections: (1) Introduction--discussing trends in the philosophy of language in regard to the role and importance of formalization, and describing and giving some history of the three fundamental notions of logical syntax, meta-language, and an "Lsemantic" definition of truth; (2) Montague--describing his work on the formalization of natural language and giving an account of Montague grammar with examples; and (3) Wittgenstein--showing how some of his ideas are relevant to the present situation of AI and the way they clash with the views of the formalist school. The course emphasized the importance of partial/inductive knowledge in our understanding of language and concluded that to handle language "formally" on a computer, it is in no way necessary to accept tenets based on formal logic. The most fruitful approaches to understanding language are precisely those not subservient to a powerful logical or semantic theory. (TL)

COURSE NOTES FOR A

TUTORIAL ON COMPUTATIONAL SEMANTICS

GIVEN AT THE

INSTITUTE FOR SEMANTIC AND COGNITIVE STUDIES
VILLA HELENEUM
6976 CASTAGNOLA
SWITZERLAND

MARCH 17 - 22, 1975

SYNOPSIS/COURSE NOTES FOR PHILOSOPHY OF LANGUAGE

Yorick Wilks

Those appalled, as I am myself, by the generality of the title will be relieved to hear right away that I am going to attempt no more than a brisk introduction followed by some detailed, though not very detailed, discussion of only two philosophers: Richard Montague and Ludwig Wittgenstein. The effect of those names may be to provoke the questions of why those particular two, or isnt that a "sublime to the ridiculous" combination. My justification will be entirely in terms of the unity that this whole tutorial is intended to present, namely the state of Computational Semantics, which means the state of work on natural language within the Artificial Intelligence paradigm. All the courses in the tutorial centre round that focus. In the case of the philosophers above, I have chosen them not so much because of the influence they have had on the subject, which is small, but because I believe that (a) the influence of Wittgenstein is almost wholly beneficial while that of Montague is largely malign. Much of what follows will be a justification of that rather sweeping judgement.

I. INTRODUCTION

Concentrating in this way means that a number of the names of the great and the good, in what is normally thought of as the philosophy of language, will not appear. To attempt to say everything is of course to say nothing. But let me venture the judgement that we can distinguish two very broad trends in the philosophy of language, as regards the role and importance of formalization. Broadly speaking, one group of philosophers have been for it, and for as much of it as possible, while the other group has been uncompromisingly against it. With a little stretching of the imagination one can reach back and assign even the Greek philosophers to one or other of these groups. It is clear for example that Aristotle was conerably more preoccupied by logic than was Plato. But, and here the assignment breaks down, it is not clear that Aristotle in his logical work was doing philosophy of language at all: for he was proposing how language should be used in order to reason correctly and in particular by menas of the permitted figures of the syllogism. He was not proposing those figures as the "CORRECT STRUCTURE OF LANGUAGE" because, of course, his logic was all expressed in a natural language, rather than in a formalization of it. Or, to put it another way, it is not clear that one can usefully talk about a formalization and its relation to the language it formalizes, until one has a formalization that is at least superficially different from the language itself.

But by the time we get to Leibniz, in the seventeenth century, the positive attitude to formalization we are discussing suddenly appears in full bloom . Leibniz was not only a formal logician, he also believed that the formalism he proposed was the real structure of ordinary language, but without its awkward ambiguities, vaguenesses, and fallacies. In his more fantastic moments he envisaged the replacement of ordinary language by this "Universal Characteristic", to the general improvement of clarity and precision. He went further: "For once missionaries are able to introduce this universal language, then will also the true religion, which stands in intimate harmony with reason, be established, and there will be as little reason to fear any apostasy in the future as to fear the renunciation of arithmetic and geometry once they have been learnt". You will see already that the formalist and attitude is not necessarily a dull and small-minded one!

By and large, the two centuries that followed saw this position sink almost without trace. The important change, for our purposes, came with the rise of formal logic at the turn of the last century. It began with the definition of propositional and general logic and the investigation of their properties. The earliest account of these calculi in English is Whitehead and Russells's Principia Mathematica in which stey were applied to the formalization of the notion of mathematical proof, but already Russell at least was setting out the ways in which this approach to logic was also, for him, a formalization of natural language. Like Leibniz, he wished to clear away what he thought of as the confusions of ordinary language. He was much exercised by the grammatical similarity of sentences like "Tigers are existent" and "Tigers are fierce" and how, in his view, this similarity had led philosophers into the error of thinking that tigers therefore had a property of existence as well as one of fierceness.

In the First order Predicate Calculus, as it is now called, the first sentence might go into some form such as Ex.Tx, to be read as "there exists some thing x such that it is a tiger", while the second might go into some form such as Ax.Tx--->F , to be read as "for all thingsx, if x is a tiger then it is fierce". The important thing here (and there are many alternative forms for these sentence codings) is that in none of them is there any predicate letter for "exists", in the way there is for tigerness or fierceness. Or to put it another way, the assertions of existence are always in the part before the dot, in what is called the quantifiers of the expression. There is never anything about existence in the body of the expression, to the right to the dot. And so, in the predicate calculus, the similarity of form between the two English sentences completely disappears.

Russell was not philosophically neutral about all this, for he believed that serious intellectual errors had followed from the "confusion" of the two grammatical forms. One classical argument about God's existence, for example, centred round the question as to whether a perfect being (i.e.God) had to have the property of existence if he was to be perfect. In Russell's view one could not reasonably talk about the "property of existence" at all once had seen that the two forms of sentence above did not both translate into "property forms" in his logic.

Wittgenstein was closely associated with Russell at the period I am describing, between 1910 and 1920, and was developing what is now thought of as his early philosophy. This was set out in a curious work called the Tractatus Logico-Philosophicuas, and I will not be discussing it in any detail here, for whan I come to Wittgenstein I want to talk only about his late philosophy. It is now faschionable to deny that there is really a strong difference between a late and an early Wittgenstein, but there can be no doubt that there appears to be this distinction. In that early work he proposed what is now called the "picture theory of truth", in which sentences in ordinary language signify because their structure reflects or exhibits, the same relation as that which holds between the things mentioned by the sentences. Thus, a logical form of fact like "catONmat" would be true if the relation between the entity symbols "cat" and "mat", and the relational symbol "ON", reflected the relation between the appropriate entities in the world. The problems for Wittgenstein's commentators (including himself) have always been (a) about what "reflected" could mean there, which might be clear if the relational symbol was and "cat" was to the left of "mat", on the page and so the sentence would be true if the real cat was indeed to the left of the real mat. However, none of the above was so clear if the relations were more complex and realistic such as And, (b) what the "entities in the world" were that the symbols referred to. It seemed clear that Wittgenstein did not mean the real objects out there in the world, like the cat and the mat themselves. He used the word "Gegenstände" and no one has ever been quite clear what these entities were to be. His theory of meaning at this point was more obscure than those of Russell, or Russells' predecessor.

Frege had had a "dualist" theory of meaning in which each word signified, if it did signify, in two ways: one way (Bedeutung), referred to some entity, and one (Sinn) to the sense of the word. The details of this distinction have preoccupied philosophers evver since, but the broad outline is clear. His famous example of "the Morning Star" and "the Evening Star" is still the best illustration: those phrases mean something different in that they have different SENSES, however it is also the case that they refer to the same entity or REFERENT. These of meaning theories need not detain us as we pass on, all that should be noted is that they are all referential , even when they talk of "sense", that is to say, whatever the status or nature of the thing that is "the meaning", it is an entity that is somehow pointed at by the word. In Frege's case the word points in two different ways to two quite different sorts of entity, that is all.

Two Ideas surfaced in Wittgenstein's early work that were to be very important in the logic of the Twenties: first the notion of significant and insignificant combination. For him the symbols for the Gegenstände could only be substituted into these picture forms of fact in certain ways and not others.

That is to say "Socrates is Mortal" reflected some relation of entities in the real world, but "Mortality is Socrates" did not, and not because those entities were not in that relation, but because the symbols "Mortality" and "Socrates" could not be substitued in that particular representation of fact at all since the combination made no sense. Another important notion was that in the theory of picture forms of fact Wittgenstein was putting forward the idea, even though hazily, that a theory of meaning required a theory of truth, because the way in which the combination of symbols reflected, or failed to reflect, a fact (i.e. was true or false) was the same thing, in some sense, as the way the form of fact made sense. It made sense only in so far as it reflected or failed to reflect a fact. In particular, he developed a very elementary theory of truth for the propositional calculus, a meht method called "truth tables", discovered independently the same time by C.S.Piece in the United States.

This notion is important for what follows so let me just set it out quickly here: the Propositional Calculus contains variables like p.q etc. that stand for the proposition espressed by any simple sentence such as:"John is happy". These simple propositions can be made into more complex one by mean of connectives NOT, AND OR and IMPLIES. Those who have attended Eugene Charniak's course will have seen these connectives already, as well as a simple proof of one expression in the calculus from others. That is, the proposition proved, ((POR Q) IMPLIES (Q OR P)) was derived by a four-step proof (see EC's notes). However, there is a quite different way of establishing the truth or falsity of that compound proposition, namely by the truth tables. Each of the connectives can be defined by a <u>table:</u>

P	Q	PIMPLIES Q	P OR Q	Q OR P	(P OR Q) IMPLIES (Q OR P)
T	T	T	T	T	T
T	F	F	T	T	T
F	T	T	T	T	T
F	F	T	F	F	T

The first column is the table for IMPLIES, the second and third for OR and the last column for the expression (P OR Q) IMPLIES (Q OR P). To the left of the vertical line we have the only four possible truth combinations of P and Q together -----were truth combination are expressed in terms of T for "true" F for "false".

The first column defines the meaning fo IMPLIES as that which is true unless its first entity is T and the second F. Similarly, the second and third columns define OR as that which is always true unless one of the constituent intems is false, i.e. it is always T except on the bottom line where both P and Q are F. Now we can construct the column for the complex expression from the columns for the simples ones. We already know that an IMPLIES expression is true except when the first entity in it is T and the second F. The first entity is the whole of (P OR Q) and we can see that it is l only F when P is F, and Q is F (the bottom line). But when that is so the right hand entity in the IMPLIES pair, namely (Q OR P), is F, hence we find ourselves writing T in every place in the column for the complex expression. Hence it is true for all possible combinations of truth values. and that is what is meant by logically true.

Thus we have stablished the truth of the expression by a completely different method, and this difference of methods is very important in what follows: let us refer to the method in Charniak's paper as proof theoretic or syntactic (it comes from the sequential relation of structures in a proof), and the second just demonstrated as semantic.

That last word is so troublesome that it cannot be introduced into these notes without a word of warning, because on its ambiguities rests much of the difficulty of our whole subject. But notice that, as I have used it there, and will use it in the next pages, it does not mean what it does in the title of the Tutorial Course. I will emphasise this by writing it "Lsemantic" when I use it in connexion with formal logic, as now.

I have introduced two of the three fundamental ideas on which the formal logic of the Twenties rest. That logic, in the hands of Tarski and Carnap in particular, is the centre of the background to this philosophy course, because the two authors to be discussed in detail represent respectively a reaction to and an extension of that logic. The three fundamental notions that I shall now turn to, are logical syntax, meta-language and a Lsemantic definition of truth.

Carnap in his "Logical Syntax of Language" developed in a systematic way the notion of "ill-formed expression" that we met earlier in discussing "Mortality is Socrates". Carnap distinguished, for any logic, its rules of formation and rules of transformation (those who think this is beginning to sound familiar should remember that Chomsky was many years later a student of Carnap). The rules of formation determined what were, and were not, well-formed expressions in the logical language so that, in the Propositional Calculus (P IMPLIES (OR Q)) was not, while (P OR Q) was, a well-formed expression. The rules of transformation then operated on those well-formed expression that were true to produce theorems in the logic. Carnap was a formalizer in the Leibnizian sense, for, to him, these distinctions applied to an ideal formalization of natural language, and in the book I referred to he tried to construct one for natural language. Carnap distinguished two types of what he called "pseudo-statement":

(1) Caesar is and

(2) Caesar is a prime number

The first was counter-syntactic in that the last word came from the wrong part-of-speech category. The second was syntactically correct but violded the rules of logical syntax. The details of all this now seem a little primitive but it was clear what he was after. In his system he also distinguished between an object- and meta-language. So, if (1) above was a syntactically correct statement in the object language (what Carnap called Language I)., then (3) Sentence (1) is syntactically correct and

(4)in sentence (1) "Caesar" occurs immediately before "is" are both statements in the metalanguage, or Language II. Carnap was far from dispassionate in all this, in that his purpose was what Russell's had been; to do away with certain kinds of sentence, and particularyl those that arose in certain kinds of philosophical writing, by showing that they violated the rules of logical syntax. The language meta-language distinction did not arise from any considerations about the structure of natural language, but was Tarski's solution, or rather part of it, to an apparently intractable problem of logic that he had inherited from Russell, who had discovered certain logical paradoxes. Tarski's statement of the problem was often in terms of the example:

THE SENTENCE IN THIS RECTANGLE IS FALSE

where any attempts to assign a truth-value to the sentence lead to trouble. Tarski thought that this problem would be solved if we only used "true" in a metalanguage and never in an object language. Thus "That John is happy is true" would be a sentence with level confusion, for the proper form would be the metalanguage sentence "'John is happy' is true", where the sentence "John is happy" is mentioned in the metalanguage sentence but not used in it.

Tarski's fundamental achievement was a theory of truth and logical consequence for formalised languages, or what in our convention we may call a Lsemantic theory.

Like many apparently revolutionary theories, Tarski's is in fact a systematisation of ideas that had existed for a long time. It used to be conventional to say that classical logic did not have a theory of truth, that it was wholly syntactic, or proof-theoretic. That is to say that Aristotle's syllogistic gave no more than forms of inference, such that one form was derivable from another, by rule, in the manner of the Propositional Calculus derivation in EC's notes that I referred to. There was, this view goes, nothing analogous to the truth table method that I described, because all syllogistic is of the general form.

PREMISE

PREMISE

CONCLUSION

such as:

Some Panthers are not Mammals

Some Mammals are not Swans
and so
Some Panthers are not Swans.

On this view, logical derivability is independent of meaning, in that it is of no importance what is put in place of the words "Panther" "Swan" and "Mammal".

However, this view is not correct, and it was known in ancient times that the last line *did not* follow from the other two, and for the following reason: suppose we replace "Panther" "Swan" and "Mammal" by "Pig" "Swine" and "Mammoth" respectively. Then we get the following inference form (where as before the line denotes the inference)

Some Pigs are not Mammoths
Some Mammoths are not Swine
Some Pigs are not Swine.

Here the premisses are true and the conclusion clearly false, so the new conclusion *cannot* follow from the new premisses, and so the old conclusion does not follow from the old premisses. This fact was known in ancient times, and by this totally non-proof theoretic method: one in which the meanings of "Pig" "Swine" and "Mammoth" were *essential* to the demonstration.

One might say, at the risk of enormous simplification, that Tarski's theory of consequence and truth is a systematic generalisation of this notion, of Leibniz's slogan "Logical truth is truth in all possible worlds" and of the truth-table notion that the logical truth of compound espressions is to be settles in terms, and only in terms, of the *truth-conditions* of the simpler propositions of which they are constructed.

One could say that the heart of Tarski's Lsemantics is his very general definition of logical consequence, and his very general definition of logical truth as a special case. The questions that then arise for tecnical logicians are how far this definition of consequence is the same as (or) "equipollent" with in the tecnical vocabulary)the proof-theoretic one and, for our purposes, how this definition of truth is to be explained. Here are Tarski's own definitions:

> Let L be any class of sentences. We replace all extralogical constants which occur in the sentences belonging to L by corresponding variables, like constants being replaced by like variables, and unlike by unlike. In this way we obtain a class L' of sentential functions. An arbitrary sequence of objects which satisfies every sentential function of the class L' will be calles a model or realization of the class L of sentences (in just this sense one usually speaks of models of an axiom system of a deductive theory). If in particular the class L consists of a single sentence X, we shall also call the model of the class L the model of the sentence.X. In terms of these concepts we can define the concept of logical consequence as follows:
>
> > The sentence X follows logically from the sentences of the class K if, and only if, every model of the class K is also a model of the sentence X.
>
> We can agree to call a class of sentences contradictory if it possesses no model. Analogously a class of sentences can be called analytical if every sequence of objects is a model of it. Both of these concepts can be related not only to classes of sentences but also to single sentences......We can also show...... that those and only those sentences are analytical which follow from every class of sentences (in particular from the empty class) and only those contradictory from which every sentence follows.

So we can see that the whole "Panthers, Mammals, Swans" argument cannot be analytic (i.e. logically true) on Tarski's definition because we found a sequence of objects (i.e. a Pig, a Mammal, and a Swan) that was not a model for it.

Notice immediately that this definition of analytic, or logically true, in terms of all models is in a clear sense, a generalisation of the notion illustrated by the truth tables earlier in terms of all distribution of trutz values. Moreover, an important notion has been introduced, under a number of different names; model, sequence, possible world - or realization, and it will appear again under the name interpretation. This notion is important in what follows when these notions are applied to natural language, and the reason the notion of "sequence" appears in Tarski's definition is because these definition were not framed for language at all, but for mathematics, so when Tarski speaks of "sequence" he is thinking essentially of a sequence of numbers, and in particular the sequence of integers. The word "sentence" in the above definitions should not confuse you into thinking that it is natural language that is being talked about.

Let us turn to another important family of Tarskian notions; those of truth definition and truth condition. The truth conditions of a sentence are, unsurprisingly, the conditions under which it is true, and the truth definition of a sentence is the specification of its truth conditions. The usual illustrative statement of the definition of truth is

"Snow is white" is true if and only if snow is white.

This is a sentence, in a meta language, of the truth condition of a sentence in an object language. It is not quite as trivial as it might appear at first sight, as can be seen by putting a German sentence in the quoted sentence space:

"Schnee ist weiss" is true if and only if snow is white.

This can now be seen to be an empirical truth about a German sentence, one that might easily have been false. Moreover, this definition cannot be generalised trivially as

"X" is true if and only if X,

because the whole point of the language-metalanguage distinction is that the first item in the sentences above is NO MORE THAN THE NAME OF THE X SYMBOL. Also,oof course, the version above in English is misleading because, as we saw, for Tarski a la uage cannot be a metalanguage for itself (i.e. we cannot have the same natural language inside and outside the quotation marks in the truth definition), because that was what gave rise to the truth paradoxes like the rectangle example. Truth definition, then, can only be in a metalanguage.

The "Snow is white*, example above can appear vacuous for other reasons too, of course, and it is perhaps a little unfair to start with it, although this is nearly always done. In an ideal Tarskian "theory of truth" sentences like the one above:

"Snow is white" is true if and only if snow is white, would appear as ultimate consequences: the final deductions from a set of truth axioms. On the way there would be more substantial looking truth conditions such as (to use an example Davidson gave for a famous tricky sentence of Bar-Hillies)

'"The box was in the pen" is true for an English speaker, at time t if and only if either the box was in the playpen before t and the circumstances surrounding x at t meet condition c (whatever that mey be yw), OR the box was in the writing pen before t and the circumstances surrounding x at t meet condition c'.'

Lastly, the "Snow" definition above is a definition of the truth conditions of the sentence "Snow is white": it is not to be considered a definition of the concept truth itself, though it would be in principle if such meta-statements were set out for all possible object sentences. That would be what might be called a definition of truth itself by extension, or by complete listing, as we might say. However, it would still be open to the charge of triviality, in that it lacks recursiveness (or calculability)

Let us see, in a tiny fragment, how recursiveness is to be put in. But for that we need two more notions, satisfaction and assignment.

Suppose we were to create a truth definition for a tiny fragment of Predicate Logic- We would have constant a:es like a, bj etc.; variables like x, y, z; predicates like F, G, R and the existential quantifier E. We would first define "sentence" or "formal expression" in the Predicate Calculus fragment, which would tell us that Rx was an expression. These rules would be what were called formation rules when we discussed Carnap: they do not tell you what is true but what is well-formed. We now encounter the one key notion of satisfaction: it is much like "true of" in ordinary language. So running is true of John ("John" satisfies the predicate "Runs", which we write Rj) when and only when John runs. We would then get what is called a recursive definition of truth for the fragment in terms of the notion of an assignment g. Let us say an expression Rx is satisfied bz an assignment g if and only if g(x) runs. An assignment is therefore a function which given a variable x which picks out an actual entity in a domain here the domain of people, and assigns it to the variable. So g(x) is a person picked out in this way, and if g(x) actually runs then Rx is satisfied by that assignment, or to put it the other way, is true-under-that-assignment. The whole thing is rounded off with a recursive definition of truth for the fragment as follows: A sentence is true if and only if it is satisfied by all assignments in the domain of persons.

Before you wonder what all this could possibly be for, let me make a few general points about it. The notion of satisfaction is the one that is like our ordinary notion of "true", and like the ascription of value T to an elementary sentence in the truth tables of the Propositional Calculus. The Tarskian notion "true" AS JUST DEFINED ABOVE is much more like the notion logic<u>ally true</u> in the Propositional Calculus: in the sense in which (A or B)IMPLIES (B OR A) was shown to be logically true. The notion of truth defined above (in "snow is white" and in the recursive definition) is normally called <u>absolute truth</u> by Tarskians, and has analogies with that of <u>analyticity</u> defined above by Tarski himself in the passage I quoted. That is to say, a sentence is analytic if <u>true for all models</u>, or all sequences of items, or, as for Leibniz, in all possible worlds.

There is something very odd about this notion of Tarski's, in particular because it does not seem to draw the common sense distinction between that something that just happens to be true, like "Lugano is in Switzerland", and something that <u>has to be true</u> like the Propositional Calculus statement above. Tarski himself was aware of this, and that, if the <u>logical constants</u> (i.e. or, implies, etc.) were chosen in non-obvious ways then any set of sentences whatever could be the <u>analytic</u> sentences of a system of logic. He accepted this consequence quite calmly. Troubles with the notion of absolute truth have caused those who want to apply Tarski's notions to natural language, like Montague, (and it is useful to remember that Tarski did not think they could be applied to natural language), to shift to the notion of "truth relative to a model" as the standard sense of "true", and not that of "true" as defined in the recursive truth definition above (truth for all assignments). This involves giving up what Tarski called "Convention T", roughly that there was a basic sense of truth independent of particular models, sequences, interpretations, and sticking only to the notion of "truth within a model"or interpretation, or what is usually called a relativistic notion of truth. It is this notion that Montague works with when discussing natural language, and this issues only in relativistic truth difinitions (giving "satisfaction conditions") rather than a full-blooded Tarskian truth definition and truth condtions. Thus, the corresponding clause to the one above, in a relative definition would contain the notion of I, an interpretation, or model, and would read:

Rx is true in interpretation I if and only if, for every assignment g, Rx is true-with-respect-to-g in I. All very heavy weather to make, you may think, but it is in fact the whole basis of Montague's semantics of natural language that I shall turn to very shortly.

Let me make two points in concluding this section:

Although Montague is the best known logician seeking to apply these general notions to natural language in detail, there is a school of logicians sharing general principles concerning the applicability of Tarksi-like theories to natural language. The thesis that they share in particular is that of the "truth theory of meaning":

namely that the meaning of a sentence is determined by its truth conditions, in the Tarskian sense of that phrase. It would be quite possible to reject Montague's detailed "Semantics for Natural Language" and still accept the general tenets of this school about meaning and truth, namely that truth conditions determine the meaning of the sentences of natural language in just the way that they can be said to do so, in Tarski's theories, for the sentences of logic and mathematics.

I shall not tackle this thesis head on in these lectures: it will be raised indirectly in what I have to say about Wittgenstein, in that his later philosophy can be thought of as in direct opposition to this view. The main figure in the group holding this thesis is Donald Davidson, and those who are interested in the gsneral principle should read Davidson's articles:

"Truth and meaning" Synthese, 1967.
"Semantics for Natura languages" in Linguaggi nella societa e nella tecnica, Edizione de Commune, Milano,1970.

and two very lucid replies

P. F. Strawson "Meaning and Truth" Oxford U. P., 1971.
(in terms of speech acts)
and
M. Dummett, "The Justification of Deduction" British Academy, 1973, in terms of a highly ingenious Wittgensteinian argument.

Secondly, the "truth-condition" approach following Tarski, should not be confused with the claims of a movement contemporary with Tarski, that of the logical positivists. They also had a thesis about the dependence of meaning on truth and it is easy to confuse it with the truth condition approach. Their principle was called the Principle of Verification, and it said "The meaning of a statement is the procedures we would carry out to establish its truth or falsehood". I shall refer to this principle when discussing Montague later, in order to claim that however wrong this principle might be (and it is wrong), it was at least serious, in a way that the modern truth condition approach is not.

Before going to the meat of the course, let me make a disclaimer at this point. I have contrasted, at some length, the formalist view of natural language with some other, undefined, approach, and we have ended up here with some more detailed description of the high point of the formalist approach, namely Montague. I would not want a reader to think at this point that, in criticising Montague, as I intend to, and to some extent espousing the views of the later Wittgenstein, I am in any sense opposing the formal analysis of natural language. Given the nature of this Institute that would be an absurd position. I am opposing what I think of as pointless formalizations of natural language, and arguing that Wittgenstein is one of the few philosophers who can provide insights into what a fruitful approach to natural language might be like.

I must tread carefully here because Wittgenstein was a foe of all attempts to apply formal logic to the analysis and understanding of natural language and, moreover, the philosopher Dreyfus, a full-time opponent of the very possibility of artificial intelligence, has made much use of Wittgenstein's arguments against formal logic in his own arguments against AI-I believe Dreyfus to be mistaken on this, and will argue that in detail later, in that I do not think that Wittgenstein's salvoes against formal logic can simply be turned round and aimed at AI- On the contrary, and this is, I think, the real philosophical importance of AI, and the reason whz it is a Wittgensteinian activity. AI has completely changed the old debate between formalists and anti-formalists. For, to handle language "formally" on a computer, it is in no way necessary to accept the tenets of Tarski, Montague or any other approach based on formal logic. On the contrary, and as you will see argued in detail in other courses, the most fruitful approaches to understanding language are precisely those not subservient to a powerful logical or semantic theory.

Thus, it is, I would claim, that AI has provided a sense to the notion of the precise manipulation of language (and anything computable must be procise of course) that is not necessarily open to the attacks of the anti-formalists (=anti-formal-logicians) like Wittgenstein. I will argue this below, but in conclusion I think it should be said that a large part of the credit for breaking the old formalist/anti-formalist opposition in a new way, must go to Chomsky

Chomsky's theory of transformational grammar, whatever its drawbacks, is certainly a precise theory of language: it also has the form of a logic, but, and here is the key originality, not the content. One way of describing transformation grammar (a non-standard but I think revealing way), is to say that Chomsky took the structure of proof-theretic logic (what Carnap you will remember called transformation rules)namely the repeated derivation of structures from other structures by means of rules of inference, but he let its content be no more than what Carnap had meant by Formation Rules, namely the separation of the well-formed from the ill-formed. Thus, in transformational grammar the inference (=transformational) rules were to apply to axioms and theorems (=kernels etc.), but to produce not new theorems but English sentences well-formed.

Thus, Chomsky had a precise system of handling language to propose (which at one stage even seemed computable), but which had no semantic definition of truth, and not even a syntactic, proof-theoretic, one either. For the notion of truth never came into the matter at all. Thus, I would claim that Chomsky's was the first concrete proposal to breach the wall between formal and anti-formal approaches, where "formal" refers to truth and not to precision. In all this AI has gone considerably further, and indeed Chomsky's paradigm still has many of the obvious drawbacks to formal approaches that I discuss here:and in particular the rigid "derivational" structure of proofs, and Chomskyan transformational derivations.

2 MONTAGUE

I argued just now that Chomsky's transformational grammar could be seen as a move to preserve the advantages of formalization, but without its "logicism". MOntague saw his task, quite clearly, as reversing that move: he explicitly began papers by saying that he intended to tackle the formalization of natural language in a way more serious than what he called "the developments emanating from MIT".

There is an initial expositional problem with MOntague's work+ most people who read it find it incomprehensible. For this reason it is far simpler to follow one of the more lucid expositions of his work, due to Barbara Partee, or Dov Gabbay. The most comprehensible of Montague's papers is almost certainly "English as a formal language" in the Linguaggi volume, refered to earlier in connexion with Davidson. That book is a good one to have, by the way, and is obtainable FREE bz writing to the Olivetti Corp., Milano, Italia.

Barbara Partee's paper is "Montague Grammar and Transformational Grammar". It will appear in Linguistic Inquiry in 1975, and until then is obtainable from her for $US 3.00 at Linguistics, Univ of Massachusetts, Amherst, Mass. USA-

Gabbay's paper, to which my account is closest, is called "Representation of the MOntague semantics as a form of the Suppes semantics" and appears in Hintikka, Suppes & Moravcsik (eds) Approaches to Natural Language. (1973)

The account I shall give here of Montague grammar is unfair to it, and it could not be sketched adequately in the time and space available. All I can try to do is give some inkling of the basic mechanism that drives it, and motivates its practitioners. The best way is to follow out the analysis of a sentence of Montague's, one that he claimed was not adequately treated by transformational grammar, and indeed could not be so treated. By that he meant sentences like " Every man loves some woman" which, to a logician have two quite different readings, and corresponding to each is a different truth condition. The "readings" of the sentence are taken to be

(i) for every man there is some woman that he loves.

(ii) there is some woman such that every man loves her.

Let us leave aside here the question as to whether or not such a sentence REALLY is ambiguous to a normal speaker of the language, that is whether or not a speaker has to have some acquaintance with the notions of formal logic in order to see that there is a reading (ii) at all. Let us also leave aside the question as to whether recent work in Generative Semantics has brought such logical ambiguities within the compass of a generative linguistics.

There are two other ingredients that must be added at this point. It is an assumption of not only MOntague, but of all those associated with the "is meaning truth" movement, that a semantics must show how the truth conditions of complex expressions are built up out of the truth conditions of their parts. That is to say, the basic scheme of analysis will be the construction of a formal semantic entity equivalent to the(truth condition of a sentence) from semantic items associated with the words that make up the sentence. Secondly, the form of the construction of this semantic entry will follow the syntactic analysis of the sentence.

In what follows I stick closely to Gabbay's version, which is much simpler than MOntague's, although Gabbay shows the formal equivalence of his schema and Montague's. Moreover, Gabbay concentrates on a single construction, whereas Montague is always so busy settling little formal points in an elegant way that it is easy to lose the main thrust of his argument. Montague's own system is set up on the basis of a categorial grammar, whereas Gabbay's is based on the more familiar, but formally equivalent notion of a phrase structure grammar. With each word category is associated both a syntatic type, that enables a phrase structure tree to be set up in a more-or-less conventional way, and a semantic type. The process consists in combining semantic types node-by-node up the syntactic tree so as to reach a single semantic item at the top. This item is then an expression of the truth-conditions of the whole sentence.

Let us start with the syntactic categories, and take only those we need for the example sentences (i) and (ii).

S = sentence

IV = intransitive verb = (run, etc.)

TV = transitive verb = (love, etc.)

CN = common noun = (man, woman,...)

Q1 = universal quantifier = (every)

Q2 = existential quantifier = (some)

we also need a derived category

NP = noun phrase.

The rules of the phrase structure grammar for our purposes are simply:

$S \longrightarrow NP + IV$

$IV \longrightarrow TV + NP$

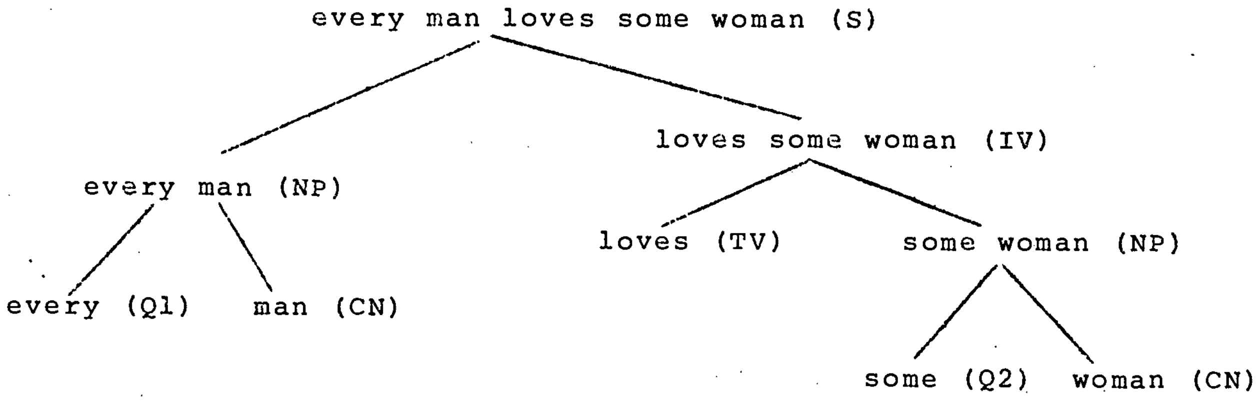
$S \longrightarrow IV + NP$

$IV \longrightarrow NP + TV$

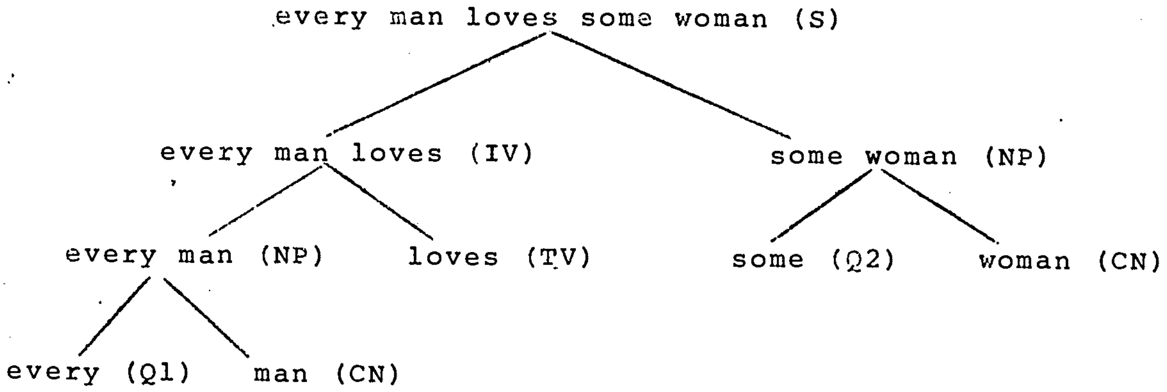
$NP \longrightarrow Q1 + CN$

$NP \longrightarrow Q2 + CN$

(i)



(ii)



Anyone having any difficulty in seeing how the two trees were obtained should realise that the first comes from taking, in order, the rules

S → NP + IV

IV → TV + NP

etc.

and the second from

S → IV + NP

IV → NP + TV

etc.

Now we turn to the semantic part, which requires the notion of I, the set of possible worlds or interpretations, and, if J is the number of instants of time, we have a set IXJ* possible worlds in all: each labelled by an ordered pair of variables (i,j). D is the domain of individuals living in these worlds. Let us think of them as people, for this sentence, and as the same people in all the worlds.. With each sentence item, whose syntactic category was given earlier, we shall now associate a semantical category giving its "meaning", which will be written as vertical bars round the word(s). Thus "John" is the name of a person, and with it we associate an element in the domain called $\|\text{John}\|$, written $\|\text{John}\|$ D. This is oversimple and we will actually treat $\|\text{John}\|$ as a set containing just John, so as the make $\|\text{John}\|$ of the same type as $\|\text{John and Mary}\|$. This is far simpler than MOntague, who treates John as a set of his properties, but the formal points will remain the same.

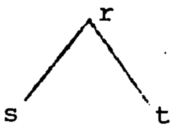
Running will be a property of people and $\|\text{run}\|$ will be a function that gives for each world (i,j) the set of those who run, i.e. $\|\text{run}\|$ i, j $\subseteq$ D. Similarly $\|\text{love}\|$ i, j $\subseteq$ DXD and is a list of who loves whom in world (i, j), i.e. a list of ordered pairs.

* "X" denotes the cross-product of two sets. If we have sets (i, i_2, i_r, i_n) and (j, j_2, J_r, J_n) then the cross-product is the set of all possible pairs like $\langle i_r\ j_r \rangle$. If one set has n things and the other has m, then there will be n x m pairs in the cross-product, where "X" means multiplied by.

There are two more sorts of rules:
one defines the properties of $\| \frown \|$ by showing the correspondence between the syntactic and semantic assignments, so we get rules, in Gabbay's numbering, where as before $\in$ is to be read is a member of, $\subseteq$ is to be read is contained in, and $\{\ |\ \}$ is to be read the set such that.

(1) $x \in IV \Rightarrow \|x\|_{ij} \subseteq D$
(5) $x \in CN \Rightarrow \|x\|_{ij} \subseteq D$
(11) $\|every\|(Do) = \{E | E \supseteq Do\}$
for some domain $Do \subseteq D$
Q1 $\Rightarrow \|every\|\ (Do) = \{E \mid E \supseteq Do\}$
Q2 $\Rightarrow \|some\|\ (Do) = \{E \mid E \cap Do \ \# \ \emptyset\}$

What we do now is to "climb up" the structured trees using another set of rules that combine these semantical objects in a predetermined way. For any node dominating two lower nodes like this:



we can construct the semantical object kr at r from the objects at kt at t and ks at s. These combination rules correspond one to one with the syntactic rules, and this correspondence is simply given (i.e. not itself constructed by a higher rule). So, we could set out two example syntactic rules that we have already encountered alongside the corresponding semantic combination rules, where, in each case, the left hand side of the syntactic rule refers to the syntactic category at node r, and the two syntactic categories on the right-hand-side refer to the categories at s and t respectively.

NP $\longrightarrow$ Q1 + CN   Kr = Kt (Ks), for each (i,j)

S $\longrightarrow$ NP + IV   for each (i,j), Kr is T if and only if (Kt   Ks) otherwise false.

To interpret these semantic combination rules you must realise that they do not assert inclusion of syntactic categories (i.e. the first semantic rule does <u>not</u> mean the application of CN to Q1 at all, but the application of the semantic object at the same node as CN to the semantic object at the same node as Q1. So Ks will be an lsemantical object corresponding to an Q1 entitz, namely a function as in rule ( ) picking out all sets in the domain that contain "all of some sort of thing", while CN will be the sort of thing in question, say, men or women. Then the result of the application will be sets containing, say, all the men in the domain.

Here are Gabbay's own applications of these rules, and others of the same form to construct the items up the trees (i) and (ii)+

*To be righthand side read as "the set of all things E such that the set overlap (intersection) of E and Do is not empty".

(a) The labels for the tree of Figure 1 are+

$\|\text{man}\| \subseteq \underline{D}$ $\|\text{every man}\| = \{\underline{E} \mid \underline{E} \supseteq \|\text{man}\|\}$

$\|\text{some woman}\| = \{E \mid \underline{E} \cap \|\text{woman}\| \neq 0\}$

$\|\text{loves some woman}\| = \{\underline{a} \mid \{r \mid (\underline{a},\underline{r}) \in \|\text{loves}\|\} \|\text{some woman}\|\}$

that is, all elements $\underline{x}$ such that there exists a $\underline{y}$ such that $\underline{x}$ loves $\underline{y}$

$\|\text{sentence}\|$ = truth if $\|\text{loves some woman}\| \in \|\text{every man}\|$,

that is, the sentence is true if for every man $\underline{x}$ there exists a woman $\underline{y}$ such that $\underline{x}$ loves $\underline{y}$.

In order to understand this labelling it is important to grasp the functional "selection" of entities that it assumes: so, for example, $\|$loves$\|$ is a set of pairs of lovers and loved, so $\{r \mid (a,r) \in \|\text{loves}\|\}$ is a set of all things in the domain that a loves.

There is really no more to the basic idea than this: namely the construction of a 1semantical object that states what it is, in set-theoretic terms, for the corresponding reading of some sentence to be true. The advantage of the Gabbay-Suppes 1semantics over the one Montague presents is not just its theoretical simplicity, but that it gets over one major fault of the Montague system;that all "logically false" sentences have the same representation. This is very counter-intuitive, for we feel fairly sure that, WHATEVER "I have proved the completeability of arithmetic" and "This is a round triangle" mean, they certainly don't mean the same. In the Gabbay system, the "meaning" of a sentence can be identified, not just with the semantical object found, but with the constructive process of assigning it up the syntactic tree. In that way Gabbay claims, two "logically false" sentences with the same 1semantical object at the top of the tree can be said to "mean something different" because of the two tree-construction processes that gave rise to them.

I have hardly set out an 1semantic theory in enough detail to justify any detailed criticism, yet nonetheless, the main outlines are there, and I think I can sketch out the form that criticism should take. (what follows is schematic, and we will I hope develop it in discussion.)

(1) One could argue that the syntax is arbitrary and unmotivated. There are indeed two syntactic readings for the sentence, which is what was wanted, but no one given two tree diagrams could guess WHICH CORRESPONDED TO WHICH READING! There will also, of course, be two readings for "Every man loves ice cream", to no particular purpose.

(2) The lsemantics is entirely reflected from the syntax and the two could not in principle diverge. This seems extraordinary, and very implausible. Consider what will happen with the Chomsky examples "John is easy to please" and "John is eager to please". (Partee and those who think like her, that transformations and lsemantics can be fused, would probably argue here that this will be cleared up in a joint system.)

(3) The assumption at every stage is that there is a molecular confrontation between language and the world. This seems plausible enough perhaps for 'John loves Mary' but wildly improbable for sentences whose meaning is explained by their inferential structure to other sentences. I will return to this in the next section, but consider how far it is from any AI, or "frames", view of meaning, on which we cannot talk about meaning independently of large structures of knowledge existing, as it were, outside the sentence examined. There is no place for that in Montague's system because meaning is to be built up only from simpler lsemantical objects, attached to the items of the sentence directly.

(4) We can contrast the triviality of the lsemantic view with the seriousness of what I earlier called the logical positivist view of meaning, that what a sentence means is the procedures we would carry out to see if it

was true. A logical positivist, faced with a difficult sentence like "God is good" might talk about what conceivable observations would be relevant to checking up on, and so giving meaning to it. But on an lsemantic view the meaning comes down in the end to some structure like (True if and only if "God" is in the class of "Good things"), indicating that truth conditions a trivialisation of a serious empirical notion.

(5) Is the notion of "truth-condition" computable? In a clear sense it is not, in that there is a possible world corresponding to every real fraction of an inch by which I am taller than Napoleon, say, and so there is at least a denumerable infinity of possible worlds in which "Wilks is taller than Napoleon" is true. Computing over them would clearly be no joke. Even were some contraction possible it is hard to see that the notion of a "class of things that run" is a useful form of manipulable information about the world. Again procedures have to be represented by static sets in lsemantics. Consider "8 is greater than 5". To establish the truth of that with computer we would do a calculation. In lsemantics we would have to search the set GREATERTHAN which CONTAINS ALL POSSIBLE PAIRS SUCH THAT ONE MEMBER IS GREATER THAN THE OTHER. Some set, some computation!

A cynic might say that, whatever the value of lsemantics as a subsequent axiomatisation, or reconstruction, of linguistic computations, it could never be a research tool; one in which important rules were established initially. In the same way, science is never done by thinking about the axioms of formal scientific systems. These notions of lsemantics are all mathematical notions and belong there. In the world of natural language, they are, in Wittgenstein's phrase, "on holiday" and cannot be expected to earn their keep.

3 WITTGENSTEIN

Wittgenstein shares one feature with Montague, that of being a "difficult" writer. There is no hope of doing more than taking a number of loosely connected topics, and giving under each a few basic quotations followed by some small smount of exposition and some remarks on its relevance to the present situation in AI or on the way it clashes, where it does, with the views of the formalist school that I have just described in some detail. This will do no sort of justice to Wittgenstein at all: each one of these topics has already been the subject of a number of articles and books. The idea is simply to give a flavour, to those unfamiliar with him, of what Wittgenstein has to offer.

Wittgenstein also had a peculiar style: his work, early and late, takes the form of a series of numbered remarks. Some of these were arranged in their present order after his death (in the early Fifties) by editors. The remarks are not themselves arranged neatly under headings, and reading Wittgenstein therefore takes the form of tracing connections through the remarks for oneself. In what follows I shall quote mostly from his Philosophical Investigations (and occasionally from his Philosophical Remarks). Both are available in German-English parallel texts. Those interested should certainly get hold of a copy of the Investigations and try the style for themselves. I will also give numbers of additional paragraphs that could be consulted under each topic. (I will use the dollar sign $ before paragraph numbers and p, as is normal, before page numbers. The page numbers are the same in English and German editions.)

The best book on Wittgenstein's later work is probably:
A. J. Kεney, Wittgenstein.

Max Black's "Wittgensteins Philosophy of Language" in his Margins of Precision is a good introductory essay. Those interested in seeing how Wittgenstein's arguments can be turned into an attack on the very notion of Artificial Intelligence, should look at:
H. Dreyfus, What Computers can't do.

There will be much I shall leave out including many of Wittgenstein's other concerns that are very close to those of AI: understanding and behavious, what it means to model the brain, whether it makes any sense to talk of "decoding the brain". I shall not discuss these major issues except in so far as they relate to language directly. Again, a whole school of modern philosophy ----the "speech act" school of Grice and Searle ---- is finding a place within modern linguistics, and also has its genesis in this same work of Wittgenstein's and his concern with notions of intention and linguistic performance. That too will have to be left out.

When reading Wittgenstein, a number of unmentioned presences have to be kept in mind at all times. The major one is Wittgenstein's early self, and his "picture theory of truth". Much of the motivation of the Philosophical Investigations was to set out why that view and its associated doctrines were wrong. Also in the background are Tarski and Carnap, who still advocate the formalist view strongly, long after Wittgenstein had given it up. Most of the views attacked in PI were held in one form or another by Tarski. The simple change in Wittgenstein was that he had ceased to believe that words had meaning chiefly becasue they pointed at things, and that sentences were true because they matched up to the world in some direct one-to-one way. He became more and more convinced that what was important about language was its "deep grammatical forms", and it was from here that the metaphor of "depth" in modern linguistics took off. Wittgenstein always resisted any actional attempt to formulate a theory of these forms, and there is no point in imagining that he would have been deliriously happy had he lived to see modern linguistics and AI as alternatives to the logical paradigm.

But I shall try to show that many of the concerns of modern AI are already there in his work, and that his line of thinking is a powerful antidote to the naive errors with which the subject is still riddled, and hence that McCarthy a leading practitioner AI was quite wrong in his judgement that "Wittgenstein set philosophy back 50 years."

Here is an epigraphic quote for all that follows:

(i) <u>Reference</u>

$ 122. A main source of our failure to understand is that we do not command a clear view of the use of our words.--Our grammar is lacking in this sort of perspicuity. A perspicuous representation produces just that understanding which consists in 'seeing connexions'. Hence the importance of finding and inventing intermediate cases.

The concept of a perspicuous representation is of fundamental significance for us. It earmarks the form of account we give, the way we look at things. (Is this a 'Weltanschauung'?)

Thesis: words do not <u>in general</u> have meaning in virtue of pointing at object in the real world (or "conceptual objects" either).

$ 35. There are, of course, what can be called "characteristic experiences" of pointing to (e.g.) the shape. For example, following the outline with one's finger or with one's eyes as one points.--But this does not happen in all cases in which I 'mean the shape' and no more does any other one characteristic process occur in all these cases.--Besides, even if something of the sort did recur in all cases, it would still depend on the circumstances--that is, on what happened before and after the pointing--whether we should say "he pointed to the shape and not to the colour".

For the words "to point to the shape", "to mean the shape", and so on, are not used in the same way as these::"to point to this book (not to that one), "to point to the chair, not to the table", and so on--Only think how differently we learn the use of the words "to point to this thing", "to point to that thing", and on the other hand "to point to the colour, not the shape", "to mean the colour", and so on.

$ 2. That philosophical concept of meaning (i.e. of meaning as pointing) has its place in a primitive idea of the way language functions. But one can also say that it is the idea of a language more primitive than ours.

$ 13. When we say: "Every word in language signifies something" we have so far said nothing whatever; unless we have explained exactly what distinction we wish to make. (It might be, of course, that we wanted to distinguish the words of language (8) from words "without meaning" such as occur in Lewis Carroll's poems, or words like "Lilliburlero" in songs.

$ 30. So one might say: the ostensive (i.e. pointing to) definition explains the use--the meaning--of the word when the overall role of the word in language is clear. Thus if I know that someone means to explain a colour-word to me the ostensive definition "That is called 'sepia'" will help me to understand the word.--

> $ 32. Someone coming into a strange country will sometimes learn the language of the inhabitants from ostensive definitions that they give him; and he will often have to guess the meaning of these definition; and will guess sometimes right, sometimes wrong.
>
> And now, I think, we can say: (those who believe in "meaning is pointing") describes the learning of human language as if the child came into a strange country and did not understand the language of the country; that is, as if it already had a language, only not this one. (See also $$11 and 27)

Comment: Wittgenstein is arguing that pointing or referring is in principle a vague activity. It can only be made clear by explaining what we are point at from within the language -- i.e. pointing assumes the whole language, except in the case of children, and the analogy with them is false ($32). As always, Wittgenstein says we could have a language based on the referential notion ($2), but it would be a language more primitive than what we call natural language. The relation of this point is the referential assumption of both Montague and many AI workers like Winograd should be obvious. In Winograd's case, it is harder to see because of the appeal to a "Procedural view of meaning". But notice that the "procedures" in Winograd all depend on the location and manipulation of some physical object, such as a block. It is not clear how a Winogradian system could function in a world that did not consist of locateable, identifiable objects, such as say, the world of newspaper articles, or these notes.

(ii) Mini languages and language games

Thesis: We can construct mini-languages obeying any rules we like, let us think of them as games. The important question is whether these games are sufficiently like the "whole game" of natural language. This question does not have a definite answer any more than this question "can one play chess without the Queen?"

Wittgenstein attributes the "pointing view of meaning" to St. Augustine and proceeds to construct a mini-language of commands and objects like "block" "clab", and colours like "red" etc.

> $ 2. Let us imagine a language for which the description given by Augustine is right. The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out;--B brings the stone which he has learnt to bring at such-and-such a call-- Conceive this as a complete primitive language.
>
> 3. Augustine, we might say, does describe a system of communication; only not everything that we call language is this system. And one has to say this in many cases where the question arises "Is this an appropriate description or not?" The answer is: "Yes, it is appropriate, but only for this narrowly circumscribed region, not for the whole of what you were claiming to describe."
>
> It is as if someone were to say: "A game consists in moving objects about on a surface according to certain rules..."--and we replied: You seem to be thinking of board games, but there are others. You can make your definition correct by expressly restricting it to those games.

Comment:

The mini-language Wittgenstein constructed should sound familiar to those of you who attended the Parsing survey course!

At this point I shall want to draw in the notion of "semantic primitive" as used in AI and linguistic systems. I shall argue that they too belong in language games and have an irreducibly linguistic character. That is to say, semantic primitives, like MAN/PHYSOB do not refer to real world objects, or Mind/brain objects either, and it is a theoretical mistake to seek to justify them in that way. They belong in a reduced language, but a language nonetheless. I shall relate this problem to how we should choose primitives and to some recent psychological results on "semantic memory".

(iii) Family resemblances and boundaries.

Thesis: The conventional notion of what a concept is,is wrong: namely, the view that a concept relates in some way to the qualities of characteristics that all things falling under the concept have. As, for example, one might claim, in a simple-minded way that everything that is an arch has such-and-such properties. Wittgenstein takes the concept of a Game and argues that one could not define a game by necessary and sufficient qualities. For any proposed necessary characteristic of being a game, Wittgenstein can think of a game that does not have the characteristic. Patience (Solitaire) for example is not competetive and so on. From this he argues that entities under a concept form a something more like a family, just as some members of a family shar characteristic X, some characteristic Y. The moral he draws is that there are not firm boundaries to concepts, nor are there to linguistic usage, or to the application of linguistic rules.

$ 69. How should we explain to someone what a game is? I imagine that we should describe games to him, and we might add: "This and similar things are called 'games'". And do we know any more about it ourselves? Is it only other people whom we cannot tell exactly what a game is?--But this is not ignorance. We do not know the boundaries because none have been drawn. To repeat, we can draw a boundary--for a special purpose. Does it take that to make the concept usable? Not at all! (Except for that special purpose.) No more than it took the definition: 1 pace = 75 cm. to make the measure of length 'one pace' usable. And if you want to say "But still, before that it wasn't an exact measure", then I reply: very well, it was an inexact one.--Though you still owe me a definition of exactness.

$ 70. "But if the concept 'game' is uncircumscribed like that, you don't really know what you mean by a 'game'".---When I give the description: "The ground was quite covered with plants"--do you want to say Idon't know what I am talking about until I can give a definition of a plant?

$71. One might say that the concept game is a concept with blurred edges.--"But is a blurred concept a concept at all?"--Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?

$ 76. If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can then be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference.

$ 84. I said that the application of a word is not everywhere bounded by rules. But what does a game look like that is everywhere bounded by rules? whose rules never let a doubt creep in, but stop up all the cracks where it might?--Can't we imagine a rule determining the application of a rule, and a doubt which it removes--and so on?

$ 80. I say "There is a chair". What if I go up to it, meaning to fetch it, and it suddenly disappears from sight?---"So it wasn't a chair, but some kind of illusion". But in a few moments we see it again and are able to touch it and so on.---"So the chair was there after all and its disappearance was some kind of illusion".---But suppose that after a time it disappears again---or seems to disappear. What are we to say now? Have you rules ready for such cases---rules saying whether one may use the word "chair" to include this kind of thing? But do we miss them when we use the word "chair"; and are we to say that we do not really attach any meaning to this word, because we are not equiped with rules for every possible application of it?

$ 88. If I tell someone "Stand roughly here"--may not this explanation work perfectly? And cannot every other one fail too?
But isn't it an inexact explanation?---Yes; why shouldn't we call it "inexact"? Only let us understand what "inexact"means. For it does not mean "unusable". And let us consider what we call "inexact".

$ 99. The sense of a sentence--one would like to saw--may, of course, leave this or that open, but the sentence must nevertheless have a definite sense. An indefinite sense--that would really not be a sense at all.--This is like: An indefinite boundary is not really a boundary at all. Here one things perhaps: If I say "I have locked the man up fast in the room--there is only one door left open"-- then

> I simply haven't locked him in at all; his being locked in is a sham. One would be inclined to say here: "You haven't done anything at all". An enclosure with a hole in it is as good as none.--But is that true?
>
> $ 100. "But still, it isn't a game, if there is some vaguness in the rules".--But does this prevent its being a game?--"Perhaps you'll call it a game, but at any rate it certainly isn't a perfect game." This means: it has impuri ties, and what I am interested in at present is the pure article.--But I want to say: we misunderstand the role of the ideal in our language. That is to say: we too should call it a game, only we are dazzled by the ideal and therefore fail to see the actual use of the word "game" clearly.
>
> $ 133. It is not our aim to refine or complete the system of rules for the use of our words in unheard-of ways.

Comment:

There are many connexions between this position and those encountered in modern AI and linguistics. Let me suggest just two for discussion.

First, can we have a serious computational semantic system until we have a self-extending one; one able to try things out, know that it had gone wrong, or re-draw built in boundaries. Secondly, can we work in this field and believe that there is a right set of rules of any sort, one to be confirmed or disconfirmed in the way a scientific hypothesis is?

(iv) The linguistic whole and confronting the world

Thesis: A language is a whole and does not confront the world sentence by sentence for the testing of its truth or falsity. The conventions of the language itself determine what are the criteria of truth and falsity in different areas of discourse.---they are different in mathematics, jokes, history, fortune cookies, advice columns, science, psychiatric interviews etc.

> $199. To understand a sentence means to understand a language. To understand a language.

Comment:

This thesis is clearly incompatible both with Wittgenstein's own early "picture theory of truth", and with any theory like Montague's, where the assumption is precisely that each sentence of a language has its truth (and its meaning) tested individually and in isolation. I shall argue in some detail that Wittgenstein's view is not at all inconsistent with a standard view of scientific truth, where sentences such as "This particle has spin 1/2" "Rats are carriers of plague" are not tested directly but belong only within large systems of inference that must be tested indirectly if at all.

(v) Logicians have a false picture of how language is.

Thesis: logicians think that language is like their favourite calculus, but they are quite wrong. Moreover, it is language itself and its use that is the standard for testing disputes that arise, not what logicians dictate.

> $81. F. P. Ramsey once emphasized in conversation with me that logic was a 'normative science'. I do not know exactly what he had in mind, but it was doubtless closely related to what only dawned on me later: namely, that in philosophy we often compare the use of words with games and calculi which have fixed rules, but cannot say that someone who is using language must

be playing such a game.---But if you say that our languages only approximate to such calculi you are standing on the very brink of a misunderstanding. For then it may look as if what we are talking about is an ideal language. As if our logic were, so to speak, a logic for a vacuum.--Whereas logic does not treat of language--or of thought--in the sense in which a natural science treats a natural phenomenon, and the most that can be said is that we construct ideal languages. But here the word "ideal" is liable to mislead, for it sounds as if these languages were better, more perfect, than our everyday language; and as if it took the logician to shew people at last what a proper sentence looked like.

All this, however, can only appear in the right light when one has attained greater clarity about the concepts of understanding, meaning, and thinking. For it will then also become clear what can lead us (and did lead me) to think that if anyone utters a sentence and means or understands it he is operating a calculus according to definite rules.

$91. But now it may come to .look as if there were something like a final analysis of our forms of language, and so a single completely resolved form of every expression. That is, as if our usual forms of expression were, essentially, unanalysed; as if there were something hidden in them that had to be brought to light. When this is done the expression is completely clarified and our problem solved.

It can also be put like this: we eliminate misunderstandings by making our expressions more exact; but now it may look as if we were moving towards a particular state, a state of complete exactness; and as if this were the real goal of our investigation.

> $101. We want to say that there can't be any vagueness in logic. The idea now absorbs us, that the ideal 'must' be found in reality. Meanwhile we do not as yet see how it occurs there, nor do we understand the nature of this "must". We think it must be in reality; for we think we already see it there.

> $115. A picture held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably.

Comment:

This thesis clearly clashes head on, not only with Montague, but also with those logicians subscribing only to predicate calculus syntax who also have strong views on its applicability to language. Notice that a Wittgensteinian is not claiming that the logicians are being *inconsistent*, as between their beliefs and the way they talk every day of their lives, any more than Phlogistian theorists were being inconsistent when they speculated while their lungs kept them alive by oxidation processes. They were simply describing phenomena they had not *examined*.

(vi) *Understanding is not a feeling*.

Thesis: We have the idea that "understanding something involves, or is associated with, a special feeling of being right. But the *tests* of our being right are quite different from the feeling.

> P. 59. (a) "Understanding a word": a state. But a mental state?--Depression, excitement, pain, are called mental states. Carry out a grammatical investigation as follows: we say
> "He was depressed the whole day".
> "He was in great excitement the whole day".
> "He has been in continuous pain since yesterday".--

We also say "Since yesterday I have understood this word". "Continuously", though?--To be sure, one can speak of an interruption of understanding. But in what cases? Compare: "When did your pains get less?" and "When did you stop understanding that word?"

$139. When someone says the word "cube" to me, for example, I know what it means. But can the whole use of the word come before my mind, when I understand it in this way?

Well, but on the other hand isn't the meaning of the word also determined by this use? And can these ways of determining meaning conflict? Can what we grasp in a flash accord with a use, fit or fail to fit it? And how can what is present to us in an instant, what comes before our mind in an instant, fit a use?

What really comes before our mind when we understand a word?--isn't it something like a picture? Can't it be a picture?

Well, suppose that a picture does come before your mind when you hear the word "cube", say the drawing of a cube. In what sense can this picture fit or fail to fit a use of the word "cube"?--Perhaps you say: "It's quite simple;--if that picture occurs to me and I point to a triangular prism for instance, and say it is a cube, then this use of the word doesn't fit the picture."--But doesn't it fit? I have purposely so chosen the example that it is quite easy to imagine a method of projection according to which the picture does fit after all.

The picture of the cube did indeed suggest a certain use to us, but it was possible for me to use it differently.

$151. But there is also this use of the word "to know": we say "Now I know!"--and similarly "Now I can do it!" and "Now I understand!"

Let us imagine the following example: A writes series of numbers down: B watches him and tries to find a law for the sequence of numbers. If he succeeds he exclaims: "Now I can go on!"---So this capacity, this understanding, is something that makes its appearance in a moment. So let us try and see what it is that makes its appearance here.--A has written down the numbers 1, 5, 11, 19, 29;at this point B says he knows how to go on. What happened here? Various things may have happened: for example, while A was slowly putting one number after another, B was occupied with trying various algebraic formulae on the numbers which had been written down. After A had written the number 19 B tried the formula $a_n = n2 + n - 1$; and the next number confirmed his hypothesis.

155. Thus what I wanted to say was: when he suddenly knew how to go on, when he understood the principle, then possibly he had a special experience--and if he is asked: "What was it? What took place when you suddenly grasped the principle?" perhaps he will describe it much as we described it above---but for us it is the circumstances under which he had such an experience that justify him in saying in such a case that he understands, that he knows how to go on.

Part II, P. 181.

Even if someone had a particular capacity only when, and only as long as he had a particular feeling, the feeling would not be the capacity.

The meaning of a word is not the experience one has in hearing or saying it, and the sense of a sentence is not a complex of such experiences.--(How do the meanings of the individual words make up the sense of the sentence "I still haven't seen him yet"?) The sentence is composed of the words, and that is enough.

There are two clear connexions between this position and our interests: first, in general AI terms, W. is making the point that it is dangerous to assess "understanding" in terms other than actual and possible performance. At this point I will draw your attention to certain current disputes discussed in my other course.

Secondly, there seems to me to be a confusion in some current work in computational semantics between how we feel about our own processes, and what an automaton must do. Consider, sense disambiguation and relate the following argument of Dreyfus' (q.v. P. 228) to Schank's argument that a proper analysis system never follows a wrong path:

> Of course, it only looks like "narrowing down" or "dis-ambiguation" to someone who approaches the problem from the computer's point of view. We shall see later that for a human being the situation is structured in terms of interrelated meanings so that the other possible meanings of a word or utterance never even have to be eliminated. They simply do not arise.

Are these positions not the same as the one Wittgenstein describes in relation to "the whole use of a word coming before the mind"? Is there a moral here about taking how we feel too seriously?

(vii) Application justifies our structures.

Thesis: The significance of a representational structure cannot be divorced from the process of its application to actual language.

Bemerkungen P. 308

> $43-5. We cannot compare a picture with reality if we cannot lay it against reality as a measuring rod and 'the rod must be in the same space as the object to be measured'

Comment

This notion of "the same space" is a difficult one, and refers back to topic (iv) that the space cannot be the physical world if the "rod" is linguistic. The notion can be taken as a plea for congruence between representational structures and language: that they exist in the same logical space and can be shown to do so. At this point I would remind you of some of my arguments about application from my other oourse.

(viii) <u>Real world knowledge and forms of life.</u>

Thesis: language understanding is not independent of very general inductive truths about our human experience.

> $142. It is only in normal cases that the use of a word is clearly prescribed; we know, are in no doubt, what to say in this or that case. The more abnormal the case, the more doubtful it becomes what we are to say. And if things were quite different from what they actually are---if there were for instance no characteristic expression of pain, of fear, of joy; if rule became exception and exception rule; or if both became phenomena of roughly equal frequency---this would make our normal language-games lose their point.---The procedure of putting a lump of cheese on a balance and fixing the price by the turn of the scale would lose its point if it frequently happened for such lumps to suddenly grow or shrink for no obvious reason. This remark will become clearer when we discuss such things as the relation of expression to feeling, and similar topics.

> II. xii. If the formation of concepts can be explained by facts of nature, should we not be interested, not in grammar, so much as in nature which forms the basis of grammar?---Our interest certainly includes the way concepts

> answer to very general facts of nature. (Such facts as usually do not strike us because of their generality.) But our interest does not revert to these possible causes of concept-formation; we're not doing natural science; nor even natural history - since we can indeed construct fictitious natural history for our purposes.

Comment:

Those who have followed other courses in the tutorial will be aware of the extent to which AI workers have emphasised the importance of partial/inductive knowledge in our understanding of language. This point has been largely overlooked by linguists, and all I am doing here is drawing attention to W's way of making the point 40 years ago.