Does Combining the Computer Analogy with Logical Semantics Produce a Theory of Pragmatics?[1]

Esa Itkonen

Autonomous linguistics is that type of linguistics which describes a given language L in such a way that the basis of an (autonomous-linguistic) description d_i is constituted by (what is intuitively known about) a set of (correct) sentences A (where $a_1,\ldots a_m$ may be either self-invented or factually observed), whereas the testing ground of d_i is constituted by a different set A'.[2] Autonomous linguistics, as here defined, allows for many different descriptions $d_1,\ldots d_n$ of L, and the only inherently autonomous-linguistic criterion for choosing one description of L over the others is _formal simplicity_ (for possible qualifications, cf. Itkonen 1978:8.4). So-called mentalistic linguistics, in turn, means to describe L in a psychologically realistic way, which means that its criterion for choosing one (mentalistic) description of L over the others is _psychological reality_. The basis of a mentalistic description d_j' may be A, just as in the case of d_i, but the testing ground of d_j' must go beyond A' to include non-intuitive or external evidence. If both the basis and the testing ground of d_j', just as those of d_i, are restricted to intuitive knowledge of (correct) sentences of L, it is meaningless to speak of d_i and d_j' as if they were exemplifications of different types of descriptions. The external evidence against which d_j' has to be tested, if it, _qua_ mentalistic description, is to be genuinely different from d_i, includes, first and foremost, evidence stemming from psycholinguistic experimentation on speech production and perception. Consequently, mentalistic linguistics, to be viable, has to be conceptually connected with psycholinguistics.[3] In order to study, within psycholinguistics, _how_ sentences are produced and perceived, one has to know _what_ sentences are, which means that psycholinguistics necessarily presupposes one or another type of autonomous linguistics. And since mentalistic linguistics is conceptually connected with psycholinguistics, it follows that mentalistic linguistics, too, presupposes autonomous linguistics.[4] - To sum up, the semantic referent of d_i qua autonomous-linguistic description is simply L (but cf. Itkonen 1978:8.4), whereas that of d'_j qua mentalistic description is L combined with part of those psychological mechanisms which make someone a speaker-hearer of L. This formulation makes it clear that the referent of d_i is an integral component, more precisely the methodologically primary component, of the referent of d'_j.

Mentalistic linguistics is still in its infancy, and it is readily discernible that the more one ascends from the concrete or lower levels like phonology to the abstract or higher levels like semantics or pragmatics, the more difficult becomes a systematic use of external evidence and, hence, the construction of mentalistic descriptions in the sense defined above. In pragmatics, in fact, external evidence can occupy at most a secondary position, as attested by the fact that all current theories of pragmatics are offsprings of Searle's (1969) theory of speech acts, which is first and foremost a philosophical theory, i.e. a theory based not upon observation/experimentation, but upon philosophical reflection. The reason for this state of affairs is not difficult to see, although it seems to have been largely overlooked. Pragmatics, or more generally the theory of language use, analyzes what it is rational to do within the normative space defined by the rules of language; that is, it analyzes what it is rational to do, given what it is correct to do. Such normative concepts as correctness and rationality (which may sometimes be difficult to distinguish from each other) cannot be established by observation/experimentation; rather, they precede, or are presupposed by, observation/experimentation (cf. Itkonen 1978:2.1 and 7.4, and 1979b, Sect. I,c). This is why external evidence is (largely) irrelevant to the theories of language use dealing with the concept of rationality, which implies that such theories may be called 'mentalistic' or 'psycholinguistic' only in a quite specific sense (for discussion, cf. Itkonen forthcoming a).

Most of today's large-scale theories of speaking and understanding are based on the computer analogy of the human mind. This analogy is taken literally in the work on artificial intelligence reviewed e.g. in Boden (1977), whereas it has a

more metaphorical function e.g. in the 'procedural semantics' of Miller & Johnson-Laird (1976). Kroy (1975) and Cooper (1978) try to model the speaker-hearer by a combination of the computer analogy and the methods of contemporary logical semantics. Cooper's program is the more ambitious one, and therefore I shall concentrate on it here. I shall first describe his approach in some detail and then assess its general significance.

Cooper accepts Popper's (1965) philosophy of science (p.51). He considers the human mind as a black box and proposes to use the standard methods of experimentation to make hypotheses about what is inside the box (p.27-29). He proposes to concentrate on the linguistic component of the mind and to offer an empirical theory of pragmatics or, more particularly, of what might be called logical pragmatics (p.1).The experiments are to be performed on rational speakers of English; a psychological state in which invalid inferences are made is not actual, because a rational person is never in such a state (p.57).[5] In other words, here as in transformational grammar, the research object is the ideal speaker-hearer, who can make inferences which no real person is capable of making (p.71-72). The ideal speaker-hearer is the last authority on which inferences conform or do not conform to the rules of English (p.139). Cooper emphasizes that the model he is constructing is not a structural model in the sense that it would replicate the factual structure of the linguistic component of the human mind. Rather, he is constructing a behavioral model: all he requires of the theoretical entities which he postulates as being inside the black box is that they produce the observable linguistic behavior (p.27).

In presenting Cooper's formalism I shall omit some technical details which are irrelevant to the general argument. The concept of an 'information automaton' is defined by Cooper with the aid of the following notions: the state set Z; the set of information inputs X_I, with members of the type $\langle x,.\rangle$; the set of test inputs X_T, with members of the type $\langle x,?\rangle$; the output set $Y = \{yes, no\}$; the state transition function d: $Z \times X_I \rightarrow Z$; the output function g: $Z \times X_T \rightarrow Y$. The concept of a 'language automaton', which makes use of the sentence set S, is defined by means of the learning operation L and of the belief relation B, which in turn are defined by means of the functions d and g. L: $Z \times S \rightarrow Z$, for instance $L(z_i,s) = d(z_i, \langle s,.\rangle) = z_j$; or, to put it less formally: to say that the machine in the state z_i learns the sentence s is to say that when it is in the state z_i and receives the input information $\langle s,.\rangle$, it moves into the state z_j. $B \subseteq Z \times S$, for instance $B(z_i,s)$ iff $g(z_i,\langle s,?\rangle) = yes$; or, to put it less formally: the machine in the state z_i believes the sentence s if and only if it, when being in the state z_i, answers 'yes' to the test input $\langle s,?\rangle$.

Cooper's black-box methodology can be illustrated as follows:

1) $s_1? \rightarrow \underset{z_0}{\square} \rightarrow$ no	The machine in the initial state z_0 answers 'no' to the test input $\langle s_1,?\rangle$, which means that it does not believe s_1.[6]
2) $s_2? \rightarrow \underset{z_0}{\square} \rightarrow$ no	Nor does it in the state z_0 believe s_2.
3) $s_1. \rightarrow \underset{z_0}{\square}$	The machine is given the information input $\langle s_1,.\rangle$.
4) $s_1? \rightarrow \underset{z_1}{\square} \rightarrow$ yes	The machine answers 'yes' to the test input $\langle s_1,?\rangle$, which means that, as a result of step 3), it has learned s_1, i.e. z_0 has changed into z_1, and that it now believes s_1; more formally: $L(z_0,s_1) = d(z_0,\langle s_1,.\rangle) = z_1$ and $B(z_1,s_1)$ because $g(z_1,\langle s_1,?\rangle) = yes$.
5) $s_2? \rightarrow \underset{z_1}{\square} \rightarrow$ yes	Because the machine believes s_2 without having received it as an information input, we retrospectively infer that in z_0 it must have believed $s_1 \supset s_2$; and since it afterwards learned, and now believes, s_1, it must, by Modus Ponens, now believe s_2.

The step 5) is crucial. It shows that the state z_1, which is meant to be a model of what (rational) human beings believe, is not just the set of sentences received as information inputs. That is to say, the machine in its initial state z_0 is not a <u>tabula rasa</u>, but contains already, i.e. <u>a priori</u>, indefinitely complex logical principles. For instance, Cooper maintains (p.29-30) that once (rational) people have received the information input 'All birds are bipeds', they answer

'yes' to the test input 'Do you know that there are no birds which are not bipeds?', which means that (rational) people are credited with the knowledge of the logical equivalence '$(x)(Fx \supset Gx) \equiv -Ex(Fx \& -Gx)$'. It turns out that all truths of standard formal logic are <u>a priori</u> constituents of the belief states. Cooper (p.62-67) conceives of his (experimental) method of data-gathering as a 'What-Do-You-Know game', where questions are put to, and answers given by, the ideal speaker-hearer. In fact, he defines (p.31) "a language automaton for a particular language [as] a behavioral model of an ideal [elsewhere: perfect] What-Do-You-Know player in that language", which means that "the input-output behavior of a language automaton for a language is indistinguishable from that of a user of the language when playing What-Do-You-Know <u>correctly</u>" (emphasis added).

The concept of state z_i is central to Cooper's undertaking. He defines it and, by the same token, the learning operation L and the belief relation B with the aid of model-theoretic semantics. In what follows I shall give a maximally simple example. The domain of discourse is constituted by the set of two individuals $A = \{1,2\}$. The language under consideration contains a single one-place predicate 'P' and two individual-names 'a' and 'b'. 's' is the variable standing for any sentence, and S is the corresponding set. There are the following four interpretation functions, which connect the individual-names with individuals and the predicate with sets:

$D_1 = \{\langle a,1\rangle, \langle b,2\rangle, \langle P,\{1,2\}\rangle\}$
$D_2 = \{$ - " - $, \langle P,\{1\}\rangle\}$
$D_3 = \{$ - " - $, \langle P,\{2\}\rangle\}$
$D_4 = \{$ - " - $, \langle P,\{\emptyset\}\rangle\}$

The model set $M = \{\langle A,D\rangle\}$; for instance the model $m_1 = \langle A,D_1\rangle$. The four models represent four different 'possible worlds'. $H \subseteq S \times M$ is the satisfaction relation; 'H(s,m)' means 's holds in m' or 'm satisfies s'. For instance, $H(Pa,m_2)$ because $D_2(a) \in D_2(P)$, i.e. $1 \in \{1\}$. $H(-s,m)$ iff not $H(s,m)$; $H(s_1 \& s_2,m)$ iff $H(s_1,m)$ and $H(s_2,m)$; the other connectives are defined accordingly. $H[s] = \{m \mid H(s,m)\}$, i.e. the set of the models in which 's' holds. The state set $Z = \{z \mid z \subseteq M, \text{ and } z \neq 0\}$. $L(z,s) = z \cap H[s]$, if $H[s] \neq 0$, otherwise undefined. $B(z,s)$ iff $z \subseteq H[s]$.

To illustrate: Let us assume that the machine is in the state $z_1 = \{m_1,m_2,m_3\}$, which means that the machine entertains m_1 (in which both 'Pa' and 'Pb' are true) and m_2 (in which 'Pa' is true and 'Pb' is false) and m_3 (in which 'Pa' is false and 'Pb' is true) as equally possible. Next the machine receives 'Pb' as an information input. We get: $L(z_1,Pb) = z_2 = z_1 \cap H[Pb] = \{m_1,m_2,m_3\} \cap \{m_1,m_3\} = \{m_1,m_3\}$. Before learning one entertains more possibilities (more 'possible worlds') than after learning; when one has learned, and thus knows, everything, one entertains only one possible world, which is identical with the actual world. As for the belief relation, not $B(z_1,Pb)$, because not $z_1 \subseteq H[Pb]$. This is correct since z_1 includes m_2 in which '-Pb' holds, and therefore the machine in z_1 cannot believe 'Pb'. However, $B(z_2,Pb)$ because $z_2 = H[Pb]$.

It is obviously unrealistic to recognize only the two cases of belief and not-belief, and Cooper in the sequel develops a formalism admitting of different degrees of belief. This refinement has no methodological importance, and therefore I shall omit discussing it here.

Now I shall proceed to an over-all assessment of the approach exemplified by Cooper (1978). In artificial-intelligence work there have been attempts to formalize commonsense beliefs in such a way that the resulting formalization could, ideally, serve as a basis for the computer simulation of actions; the written program shows how (descriptions of) actions are explicitly generated (cf. Schank & Abelson 1977). This work, which has obvious similarities with explicit generation of (descriptions of) sentences within the transformational-generative tradition, is time-consuming and may seem tedious. Therefore it is comprehensible that there is a temptation to try a short-cut: to keep the computer analogy and to solve the question of belief systems at one stroke, namely by borrowing the relevant concepts from formal logic, more precisely from model-theoretic semantics. Thus Kroy (1975), for instance, simply postulates the contemporary epistemic logic as part of the human mind and intends this as a contribution to solving the not-just-logical problems of knowledge and belief. Cooper acts analogously in his attempt at constructing a theory of

pragmatics. This approach is not without merit; for one thing, it shows that formal logic is not as far removed from the human realities as the logicians themselves are generally inclined to think (cf. Itkonen 1978:2.6). However, I believe that this logicist approach to pragmatics contains serious flaws, both with respect to its methodological self-understanding and with respect to what it can reasonably hope to achieve. I shall concentrate here on three specific points.

<u>Empiricalness</u>. Cooper's view of what kind of research he is conducting rests on a complete misunderstanding. The nature of this misunderstanding becomes evident e.g. in the passage where he claims (p.87) that "empirical tests have shown belief in the [premises] to be accompanied invariably by belief in the [conclusion] among rational speakers of English". Cooper nowhere provides an independent criterion of rationality. On the contrary, from what he says elsewhere (e.g. p.73), it is clear that if someone fails the test, i.e. if he fails to recognize the principles of formal logic (even where these are in fact quite difficult to understand), he is not a rational speaker of English.

Cooper succumbs here to the same positivistic fallacy whose pervasive influence on contemporary theoretical linguistics I have analyzed in several of my writings. To be sure, he cannot help realizing that what he is doing does not at least <u>look</u> like observation/experimentation as carried out in physics. To explain away this discrepancy between 'logico-linguistics' and physics, he resorts to two standard arguments. First, he claims (p.89) that the paradigmatic situation of grammar-writing is the one in which the grammarian, aided by informants, is writing the grammar of a language unknown to him; it is only an accidental fact that he, Cooper, happens to describe a language which he knows perfectly, and thus acts, in fact, as his own informant. Always when I encounter this argument, offered by linguists who have never worked with informants or described an unfamiliar language,[7] I am amazed that people can bring themselves to make that which never happens a rule and that which always happens an exception. Second, Cooper claims (p.68-69) that the normatively binding rules of speaking and, in particular, of inferring are 'idealizations' in precisely the same sense as the idealizations used in physics; but he does not even try to show what, precisely, this supposed analogy consists in. On the basis of such an inexact use of terms like 'idealization' or 'theoretical concept' it is possible to 'prove' anything, for instance that physics and philosophy are entirely similar (cf. Itkonen 1978:7.2).[8]

It is good to point out that there is no reason why the 'What-Do-You-Know game' could not be used as a genuinely experimental device. But then it is certain that the results of the corresponding experiments would exhibit the same kind of variation as is commonly found in experimental studies on the psychology of logic (cf. Osherson 1975).

Black box. Cooper's black-box methodology is entirely superfluous because he, even before starting to play the 'What-Do-You-Know game' with himself, knows quite precisely what will be inside the box: he will place in it all standard principles of formal logic, and a player who fails to play the game in accordance with such principles is shrugged off as irrational. This conclusion is not affected by the fact that in his case study on 'if - then' (p.158-211) Cooper genuinely tries to go beyond standard formal logic and to establish something approximating 'natural logic'. What he is doing is still to analyze his own intuitive knowledge of valid inferring in English.

Pragmatics. As a model of speaker-hearers Cooper's language automaton is so abstract as to virtually lack any factual content. The automaton does not allow for people changing their minds or learning things contrary to what they have learned before, and it is deterministic, which means that all people should learn the same thing in the same situation. These remarks may be countered by pointing out that the automaton does not describe real speakers, but the ideal speaker-hearer. I think, however, that this kind of 'defense' would be fatal to the entire approach. The only acceptable defense would be to admit the gap between the theory and the facts, and to promise to narrow it in the future.[9] Moreover, the state transition function $d(z_i,\langle s,.\rangle) = z_j$ is in principle defined for all (successions of) sentences s of English, but this is obviously impossible. It would have been interesting to see whether the function can be sensibly

defined even for a small number of actual sentences, but on this Copper is silent. What he offers is a narrow insight into the formal behavior of the ideal speaker-hearer.

A theory of pragmatics has to account for the intentions which 'cause' the speaker's behavior and are hopefully recovered by the hearer.[10] In Cooper's theory what the automaton believes is defined in terms of what it answers to the test inputs inquiring about its beliefs; and these answers are supposedly to be construed as being caused by corresponding intentions. The entire verbal behavior consists of two words, i.e. 'yes' and 'no'. It should be obvious that this is a rather one-sided conception of language use, even if one is willing to confine one's attention to the logical aspects of language. A more full-blooded version of logical pragmatics is offered e.g. by Lorenzen's and Lorenz's dialogical or game-theoretical logic, in which the logical connectives and quantifiers as well as the concepts of empirical and logical truth are defined in terms of attacks upon, and defenses of, claims made by the proponent and his opponent (cf. Itkonen 1978:2.6). As far as empirical claims are concerned, Cooper is able to deal only with the speaker's subjective beliefs: if he learns a de facto false sentence, he will believe it for ever. By contrast, the dialogical approach is able to distinguish between subjective belief and objective truth: the speaker's belief is represented by the sentence he is willing to defend, but if he does not succeed in doing so, it has been shown to be false. - There is a built-in limit upon the capacity of any logical system to model human rationality. True rationality merges with creativity, which means that it is (often) unpredictable;[11] but the goal of logical modelling is the 'mechanization of deductive (or inductive) reasoning', to borrow Robinson's (1979) term.

Cooper's claim to have constructed a theory of pragmatics ultimately rests on the analogy between his language automaton and a human speaker: both have an 'inside' and an 'outside', both receive 'stimuli' from their environment and 'behave' in response to them. It should be noted, however, that if nothing more is required for the analogy to be established, then any input-output machine will qualify as a model of human beings. This is true, for instance, of the standard type of Turing machine, which is defined by the output function R, the movement function D, and the state transition function Q, as exemplified by $R(q_i,s_j) = s_{ij}$, $D(q_i,s_j) = d_{ij}$, and $Q(q_i,s_j) = q_{ij}$. To put it less formally: When the machine is in the state q_i and is reading the symbol s_j, which has the status of an input, three things happen. First, the machine prints a new symbol s_{ij}, which has the status of an output. Second, it makes a move d_{ij} either to the left or to the right. Third, it enters a new state q_{ij}. It is well known, however, that Turing machines are primarily interpretable as syntactic devices which describe a language by recognizing its well-formed strings. The illusoriness of Turing machines as behavioral models is shown by the fact that both the arithmetization and the axiomatization of Turing machines literally eliminate the distinction between the inside and the outside (cf. Minsky 1967:10.1 and 12.6); the concept of a machine and, with it, the similarity with human beings disappear in the process. Consequently, additional constraints have to be imposed upon machines if they are to qualify as genuine models of human beings. What these constraints are, precisely, is still an open question.

NOTES

1) This paper could be seen as an elaboration of the last paragraph of Itkonen (1977).

2) For simplicity, I define autonomous linguistics as description of (properties of and relations between) sentences, but it is clear that the data of autonomous linguistics has to include suprasentential units, i.e. texts, as well. For a definition of the concepts in question, cf. Itkonen (1979a).

3) Personally I am, in fact, convinced that there is no mentalistic linguistics over and above psycholinguistics: psychology of language use contains the psychology of language. But in order not to unnecessarily complicate the issue, I shall maintain here the more traditional distinction between nonmentalistic grammar and mentalistic grammar.

4) This conclusion, opposed e.g. by Derwing (1979), has been argued for by Itkonen (1974: VI,4 and VII,4) and (1978:7.4 and 8.4) as well as by Kac (1974) and (1979).

5) Elsewhere, e.g. p.69, Cooper uses the term 'actual' in a more customary sense.

6) 'It does not believe s' means 'It is not the case that it believes s', not 'It believes not-s'.

7) I do not deny, of course, that linguists sometimes describe languages which they know less than perfectly. I myself have described Merovingian Latin (cf. Itkonen 1978a), which, because of the nature of the existing texts, no one knows perfectly. But his is precisely why I know that describing an unfamiliar language (e.g. Merovingian Latin) is <u>different</u> from describing a familiar language (for me, Finnish). Linguists who have never described unfamiliar languages perhaps cannot as much as suspect the existence of this difference.

8) It is interesting to note that in contradistinction to the great majority of contemporary linguists, the students of artificial intelligence are generally aware of the fact that constructing models of rational (linguistic) behavior requires an inherently nonpositivistic methodology; cf. Boden (1977: 403-4) and Sloman (1978:63-64).

9) If the concept of 'ideal speaker-hearer' is too far removed from real speaker-hearers, it loses its justification and becomes meaningless. The problem is, of course, to decide when the idealization involved ceases to be beneficial and becomes detrimental. Making this decision may be facilitated by various types of consideration. For instance, I believe that the justification of the concept of 'ideal speaker-hearer', as employed within the transformational-generative tradition, diminishes, once it is seen that this concept is indistinguishable from the historically earlier concept of 'extended axiomatic system', which was invented for the description of <u>artificial</u> languages; cf. Itkonen (1976).

10) For simplicity I use this standard, individualistic terminology, although it is in my opinion seriously misleading insofar as it conceals the inherently social nature of understanding. Intentions do not exist outside the social forms into which they are codified (although there are differences of degree between more standard forms and less standard ones), and the understanding of intentions does not exist outside the social criteria for distinguishing understanding from misunderstanding; for a criticism of the individualistic, neo-Cartesian philosophy of psycholinguistics, cf. Itkonen (forthcoming b, chap.4).

11) If linguistic change is viewed as a type of rational and therefore unpredictable behavior, the resulting notion of diachronic-linguistic explanation is bound to differ from the traditional one; cf. Itkonen (1978b) and (forthcoming a).

REFERENCES

Boden, Margaret (1977): Artificial intelligence and natural man, Hassocks: Harvester Press.

Cooper, William (1978): Foundations of logico-linguistics, Dordrecht-Holland: Reidel.

Derwing, Bruce (1979): 'Against autonomous linguistics', in Perry (1979).

Itkonen, Esa (1974): Linguistics and metascience, Kokemäki: Risteen kirjapaino.

- " - (1976): 'The use and misuse of the principle of axiomatics in linguistics,' Lingua.

- " - (1977): 'Causal models in linguistics', Publications of the Linguistic Association of Finland,1, Turku.

- " - (1978): Grammatical theory and metascience, Amsterdam: Benjamins.

- " - (1978a): 'The significance of Merovingian Latin to linguistic theory', in Four studies on classical languages, Publications of the General Linguistics Department of the University of Helsinki, 5.

- " - (1978b): 'Short-term and long-term teleology in linguistic change', Publications of the Linguistic Association of Finland, 2; to be reprinted in P. Maher (ed.): Papers from the IIIrd international conference on historical linguistics, Amsterdam: Benjamins.

- " - (1979a): 'Distinguishing between sentence and text' in J. Petöfi (ed.): Sentence vs. text, Hamburg: Buske.

- " - (1979b): 'Qualitative vs. quantitative analysis in linguistics', in Perry (1979).

- " - (forthcoming a): 'Rationality as an explanatory concept in linguistics', in H. Geckeler et al.: Festschrift für Eugenio Coseriu, Tübingen: Narr.

- " - (forthcoming b): Causality in linguistics.

Kac, Michael (1974): 'Autonomous linguistics and psycholinguistics', Minnesota working papers in linguistics and philosophy of language, 2.

- " - (1979): 'In defense of autonomous linguistics', Lingua.

Kroy, Moshe (1975): Mentalism and modal logic, Frankfurt a/M: Athenäum.

Miller, George, and Philip Johnson-Laird (1976): Language and perception, London: Cambridge University Press.

Minsky, Marvin (1967): Computation. Finite and infinite machines, London: Prentice-Hall.

Osherson, Daniel (1975): Logical abilities in children, vol. 3: Reasoning in adolescence: deductive inference, Hillsdale: Lawrence Erlbaum Associates.

Perry, Thomas (1979): Evidence and argumentation in linguistics, Berlin: de Gruyter.

Popper, Karl (1934/1965W: The logic of scientific discovery, New York: Harper & Row.

Robinson, J.A. (1979): Logic: form and function, Edinburgh: Edinburgh University Press.

Schank, Roger, and Robert Abelson (1977): Scripts, plans, goals, and understanding, Hillsdale: Lawrence Erlbaum Associates.

Searle, John (1969): Speech acts, London: Cambridge University Press.

Sloman, Aaron (1978): The computer revolution in philosophy, Hassocks: Harvester Press.