GEOFFREY SAMPSON

Natural Language and the Paradox of the Liar*

A logistic system is called 'consistent' if none of its well-formed formulae (wffs) is such that both itself and its own contradictory are theorems of the system. If we choose to think of a natural language, say English, as a logistic system, then we may ask whether it is a consistent system. Since we can express the paradox of the Liar in natural language — that is, since we can utter an English sentence such as

A. What I am now saying is false

it may appear *prima facie* that natural language is an inconsistent system. For inconsistency to follow from the expressibility of *A* it is of course necessary not only to choose to consider the sentences of English as wffs of a logistic system, but also to postulate axioms and rules of inference for the system. Suppose we postulate 'commonsense' rules of reiteration and disjunction elimination and the 'commonsense' axiom schema '*X* is true or *X* is false', together with rules for the introduction and elimination of the phrase 'is true' according to which '*X*' entails and is entailed by '*X* is true'. Then, if we admit the existence of a wff *a* of the form '*a* is false' (so that to utter '*a*' is to make the same statement as is made by uttering '*a* is false'), there is no difficulty in writing out proofs of both '*a* is true' and '*a* is false' (which are presumably contradictories). *A*, or some sentence obtained from *A* by replacing the deictics *I* and *now* by proper names or paraphrases, and perhaps by substituting *write* or *utter*

* I acknowledge with gratitude the helpful comments of L. Jonathan Cohen, Yehoshua Bar-Hillel, Gilbert Ryle, and Carl Kordig. Responsibility for shortcomings is mine alone.

for *say*, may be claimed to have the properties required for *a*. All this is, of course, very well known.

The problem of the consistency of natural languages is created by the decision to think of natural languages as logistic systems. Bar-Hillel (1966: 392) suggests that this approach was first taken by logicians of the Polish school in 1919. Since then the issue has been discussed in numerous publications, although Bar-Hillel (393) points out that only logicians rather than linguists have dealt with it. Recent developments in linguistic theory tend to suggest that the analogy between natural language(s) and artificial logistic systems is a much closer and more natural one than might previously have been thought. So perhaps this is an appropriate juncture for a linguist to offer a contribution towards solving the consistency problem.

Until recently the accepted version of linguistic theory (cf. Chomsky, 1965) held that the abstract term 'sentence' stood for a triple of objects: a 'surface structure', which was in a certain sense available to observation, a 'deep structure', and a 'semantic interpretation' — or, to use a less tendentious term, 'semantic representation'. Given a language L, a particular surface structure of L determined a set of one or more deep structures of L, and a deep structure of L determined a set of one or more semantic representations of L. The relation between the surface structures and deep structures of a language was expressed in a set of 'syntactic transformations', and the relation between the deep structures and semantic representations of the language was expressed in a set of 'projection rules'.

The concept 'semantic representation' deserves some enlargement. Suppose the language L consists of the (infinite) set of sentences $\{S_1, S_2, \ldots\}$. Then to claim that the set of formulae $\{\sigma_1, \sigma_2, \ldots\}$ are semantic representations of $\{S_1, S_2, \ldots\}$ respectively is to claim that there exists some finite set of formal rules of inference by which $\sigma_i \vdash \sigma_j$ just when speakers of L consider that the statement made by uttering S_i implies the truth of S_j, $\sigma_k \vdash O$ just when speakers of L consider the statement made by uttering S_k to be a contradiction in terms, $\vdash \sigma_l$ just when speakers of L consider the statement made by uttering S_l to be analytically true, and so forth. Of course, enormous problems arise if the rules of inference are required to apply to lexical items — does the statement *John is a man* imply the truth of *John is rational*? But it is possible to think in terms of 'restricted semantic representations', where the entailments are required to succeed only if the relations of implication, contradiction, and so on between the corresponding sentences hold other than by virtue of the

senses of the lexical items composing them. Thus, of Quine's sentences *No unmarried man is married* and *No bachelor is married* (Quine, 1953: 22-23), the restricted semantic representation of the first, but not that of the second, is required to be a theorem of the system. For further discussion, cf. Sampson, forthcoming.

The research strategy for discovering syntactic transformations, and hence for discovering the deep structures corresponding to given surface structures in a given language, is rather well understood and agreed on. Furthermore it is empirical: although intuitions of any type, including introspection about the meanings of sentences, are permissible in forming hypotheses about syntactic transformations and deep structure, the only evidence admitted as supporting or refuting such hypotheses is the tangible evidence of surface structure — cf. Postal (1970: 102ff), Sampson (1970: 268-69). On the other hand, it has been by no means clear from the literature (e.g., Katz and Fodor, 1963; Katz, 1966, 1967) either what form semantic representations are to take, or what sort of evidence is relevant for or against hypotheses about the membership of the set of projection rules.

But now a movement grouped under the banner 'Generative Semantics' has appeared (cf. McCawley, 1968; Lakoff, 1970; Postal, 1970: 98ff), which points out: first, that neither empirical evidence nor intuitive argument has been offered for the assumption that projection rules, distinct from syntactic transformations, exist in natural languages; and, secondly, that deep structures are already approaching more and more closely to our intuitive conception of what semantic representations (at least of the restricted kind) should look like, so that inductively it becomes increasingly likely that the remaining discrepancies between restricted semantic representations of sentences and their deep structures will vanish as our knowledge of the syntactic transformations of English and other languages becomes complete. Postal (1970) even gives an argument suggesting that syntactic research may produce nonrestricted semantic representations; but nothing in this essay depends on that.

The generative-semantic hypothesis is highly relevant for the view of natural languages as logistic systems. Previous writers who have compared the two phenomena have not, of course, been able simply to treat English (e.g.) surface structures, i.e., strings of words or morphemes, as the wffs of a logistic system; it would appear impossible to frame a comprehensive but finite set of rules of inference for this set of wffs. (Think, for instance, of the lack of generality in a rule permitting the derivation of the string *Bill is believe -ed by Mary to be old -er than Alex*

from the string *Mary does not believe that Alex is as old as Bill.*) So it has been necessary to invent logistic systems that are in themselves artificial, and for which satisfactory rules of inference can be formulated, but that aim to reflect all the semantic functions that can be found in some natural language. (For an extended example, cf. Reichenbach, 1947: ch. 7.) The correlation of a particular wff of the logistic system with a particular sentence of the natural language appeals solely to the reader's introspective knowledge of the meaning of the natural-language sentence; where different individuals disagree about the appropriate way of representing some natural-language sentence-type in symbolic form, there is no objective criterion for resolving their disagreement. But if the generative-semantic hypothesis is right, linguistics provides us with a method independent of introspection for obtaining a set of wffs corresponding to the set of sentences of natural language, for which rules of inference can be framed to achieve the desired results.

(Throughout this paper I discuss facts about English in terms implying that they may be relevant to natural language in general. This is because another prediction in linguistics, in this case not peculiar to the generative semanticists, is that the system of deep structures will turn out to be identical for all natural languages, other than with respect to the set of lexical items. Bar-Hillel characterizes the Polish logicians as linguistically unsophisticated because they "almost creat[ed] the impression that when talking about our everyday affairs in English, Polish or German we were using different dialects of the same 'language of everyday speech'" [1966: 393]; if the universality prediction is fulfilled, this could be a quite appropriate way of describing the situation.)

If generative semantics is right, it will not be possible to dodge the Liar issue by saying that the notion of consistency is inapplicable to natural languages because they are vague. In this paper I shall assume the hypothesis. Thus I assume that the question of the consistency of natural language is a real one, and that the existence of a wff of the *a* sort (cf. p. 305) would suffice to prove natural language inconsistent. I shall first examine some earlier discussions of the problem, and then offer my own attempt to solve it. I should make it clear at the outset that I do not claim to prove that natural language is consistent, but only to show that there is no threat to its consistency from sentences such as *A*. I shall not attempt an exhaustive survey of the literature on the question, but shall simply select some works that seem particularly appropriate as jumping-off points for my own argument.

One early answer to the problem posed by the utterability of *A* was

given by Stroll (1954). He felt that natural language is indeed inconsistent, but that this does not matter in practice since, as a matter of pragmatics, sentences like *A* will never be useful: the set of sentences in practical use will never lead to contradiction. This position seems tenable (although I shall argue below, p. 314, that it entails an unfortunate consequence); but many of us will remain dissatisfied. We intuitively want to find that the whole set of grammatical, meaningful natural-language sentences form a consistent system, although some of them, e.g., tautologies like *All cats are cats*, have no practical use.

Ziff (1960: 138) claims that natural language is consistent. He solves the difficulty of *A* by regarding it as a 'deviant' sentence rather than a well-formed sentence of English. A 'deviant' sentence for Ziff is one that fails to conform to the 'regularities' of English usage. In his discussion of deviance (30-34) he distinguishes two subtypes: semantic incompatibility (as in the phrase *green idea*), and 'ungrammaticalness' or syntactic deviance (as in **an apple good* instead of *a good apple*). But Ziff does not suggest that *A* violates any syntactic or semantic regularity that can be established independently of considerations of logical consistency; indeed, as we shall see, he seems to assert that *A* does not violate such regularities. His 'proof' that *A* is deviant is simply that: it is a regularity of English that a sentence that is actually used (and hence nondeviant) 'generally ... does not lead to a contradiction'; *A* does lead to a contradiction, *ergo* *A* is deviant.

In the first place this is no kind of proof: that the general run of sentences in actual use do not lead to contradiction is not incompatible with the existence of exceptional sentences in use that do lead to contradiction, and furthermore Ziff assumes that all nondeviant sentences are in common use, which seems unreasonable. (cf. the remark frequently made by theoretical linguists that most sentences one actually hears or reads were never heard or read by one before.) This mistake seems to spring from Ziff's unwillingness to recognize the existence of grammatical rules (in the linguistic sense), as distinct from generalizations about surface structure ('regularities'). (NB: the linguistic use of the word 'rule' differs from the use common among philosophers, which stresses the breachability of rules. For a discussion of 'rules' in linguistics, see, e.g., Chomsky, 1961.)

But the important point to understand about Ziff's argument is that by deriving the deviance of *A* from the paradoxicality of *A*, Ziff is effectively setting up a third type of deviance. Although natural language is consistent, it is so for a special reason. In an artificial logistic system there is a set of formation rules (rules for enumerating wffs) that do not mention

consistency, and a set of rules of inference, and the two sets jointly determine the consistency or otherwise of the system. But in the case of natural language, according to Ziff, as well as the ordinary (syntactic and semantic) formation rules ('regularities') for enumerating nondeviant sentences, and the rules of inference, there is a special extra formation rule, which states that any formula that would count as a nondeviant sentence by the other formation rules, but in conjunction with the rules of inference would yield a contradiction, (e.g., *A*), is after all not a nondeviant sentence. "If in a logistic system I come across a contradiction, I cross out the system. But if in my language I find a contradiction, I cross out the contradiction" (*loc. cit.*). For Ziff, natural language as a logical system is consistent by definition.

Miss Anscombe (1963: 293) characterizes such an approach to the problem as an "avoidance of honest toil". I believe Ziff's view is one that can rationally be held: it might be that although language would otherwise be inconsistent, we simply prevent ourselves from stating (or, presumably, thinking) any proposition that could lead to contradiction. This position might seem psychologically implausible, as it is not clear how persons would determine whether or not a given proposition led to contradiction, and it would be philosophically unenlightening, since natural language would be consistent but trivially so; but it is tenable.

Both Stroll and Ziff deal with the problem of *A* by ruling it out of the set of well-formed sentences under consideration, whether this is the set in practical use or the set theoretically available; but in both cases they give no independent reason for ruling it out other than the fact that it leads to paradox. In the latest of his own contributions to the problem (1966), Bar-Hillel makes a similar view even more explicit. He draws the distinction between 'sentence' and 'statement' (statements, rather than sentences, being the bearers of truth-values), and between 'uttering a declarative sentence' and 'making a statement'; and he wishes to describe the case of someone uttering a sentence such as *A* as a situation where a declarative sentence is uttered but no statement is made. He goes on to say (396):

> Some people might find it disturbing that, in [a case of someone uttering an *A*-type sentence], there was no more direct way to determine that no statement was made by the utterance in question, other than by deriving a contradiction from the contrary assumption. ... However, I would insist that recognising the essentially non-recursive character of the notion of statement in natural languages, i.e., the non-existence of a mechanical procedure for deciding whether, given an utterance, a statement was made by it, is just another part of the price that has to be paid for keeping our natural languages consistent.

Whether natural languages are consistent should not be asked as an empirical question but decided *a priori*, Bar-Hillel believes (394). The cost of making such an *a priori* decision is that it follows that there is no decision procedure for determining whether a given act of uttering a declarative sentence makes a statement or not; but for Bar-Hillel "the only really disturbing fact about [this] is that so many people still find it disturbing" (396-97). He also claims (394) that this cost has to be paid anyway, for independent reasons; but the paragraph on p. 396 that purports to validate this claim appears to me only to suggest that the distinction between statement and utterance is independently necessary — which I agree — not that the notion of statement is necessarily nonrecursive.

I agree that the utterance of *A* is a case of uttering a sentence but not making a statement, so that the contradiction does not arise: in the discussion on my p. 305, the *X* of the axiom schema ranges over statements, not sentences, so that one may not substitute *A* but only the statement made by uttering *A* for *X* — if there is no such statement, the proof fails and no paradox occurs. But I find it highly disturbing to suppose that someone who hears a sentence addressed to him and recognizes the words composing it and its syntactic structure, might have no procedure available for determining whether or not a statement has been made by the utterance of the sentence. I do not see in that case how natural language could function as the useful tool we know it to be.

I still wish to find natural language consistent as an empirical fact, rather than declaring it to be so by *a priori* decision. To do this it will be necessary to find a principle according to which the utterance of *A* fails to make a statement, and which can be validated independently of considerations of contradiction. Ziff explicitly denies that this is possible (1960: 136). He suggests four features of sentence *A* as candidates for the cause of its oddity: the predicate *(is) false*, the self-referent subject *what I am now saying*, the reference to the first person *I (am)*, and the time identification of *am now ...ing*; and he rules them all out by stating that they can each occur in other, perfectly unexceptionable sentences.

The logician will be predisposed to assume that the trouble with *A* DOES have something to do with self-reference. Yet Ziff assures us that the self-reference of *What I am now saying* is in no way objectionable, except in this particular environment. He supports this by pointing out that the sentence:

B. What I am now saying is not to be repeated.

is not paradoxical. I certainly agree that *B* could occur in natural usage,

and that it would not lead to paradox. But would it be self-referential? Surely the only situation in which *B* could occur in natural usage would be as a parenthetical insertion in or addition to a set of other statements. (For the relevance of 'natural usage', see below, pp. 314ff.) And in such a situation, where the speaker uttered sentences asserting propositions p_1, p_2, ..., p_n, and also uttered sentence *B*, the proposition p_B that any normal hearer would understand *B* to assert would be that the set of propositions $\{p_1, p_2, \ldots, p_n\}$ were not to be repeated: it would NOT be that the set $\{p_1, p_2, \ldots, p_n, p_B\}$ were not to be repeated. And it seems clear that if a speaker uttered *B* preceded and followed by silence, a normal hearer would react by blank bewilderment, or by saying "Well, go on, say it" or the like: only a logician would consider that the speaker had made an amusingly trivial assertion.

Why is it that the utterance of *B* will be understood to ban the repetition only of the propositions other than p_B? One might suggest that in practice it would be pointless to ban the repetition of p_B (whereas there may be a good reason for banning the repetition of $\{p_1, p_2, \ldots, p_n\}$), and that the normal hearer deduces by common sense that the phrase must therefore have been intended to cover just $\{p_1, p_2, \ldots, p_n\}$. But it may be quite sensible to say

C. What I am saying now is not to be repeated; do not even pass on the fact that I have pledged you to secrecy thus.

C could be used to instruct someone who is going to face a press conference, say, to reply, "No comment" to various questions, rather than saying "My informant told me that I was not to give the answer to that question". But now notice that the meaning that *C* actually DOES have seems to be exactly the meaning that *B* WOULD have if the subject of *B* could refer to p_B. If, as I believe, it is true that in practice *B* would never be understood as meaning the same as *C* (except by a logician), then this fact cannot be explained pragmatically, in terms of sentences being interpreted only in ways that seem practically reasonable, and it must instead be explained by some definite rule of language.

I believe that the rule in question is that self-reference does not occur in natural language, in the sense that no referring expression will be taken to refer to the proposition expressed by the sentence within which it is embedded — whether or not contradiction would arise if the expression were to be taken as self-referent in this sense. Let me be more precise. The constituents of sentences that refer — Strawson's "uniquely referring expressions" (Strawson, 1950) — are constituents called by linguists

'noun-phrases' or NPs. (The term is a technical one; not all NPs contain nouns, nor are NPs necessarily more than one word long, thus both italicized constituents in "He pities *the poor*" are NPs. All referring expressions are NPs, but not *vice versa* — e.g., *a car* in *He didn't buy a car* is not a referring expression.) The words composing a referring NP function so as to cut away the members of a universe of entities, leaving a single entity, which is what the NP refers to. Thus in a sentence such as

D. I visited *the building in London which the Queen lives in*, and photographed *it*.

the first italicized NP rules out all entities that are not buildings, and in London, and inhabited by the Queen, so that it will succeed in referring no matter how large the universe of entities. The second NP, on the other hand, will succeed in referring only so long as the universe contains but a single inanimate entity, and ambiguity will occur if the hearer has reason to believe the speaker might be referring to some object within the palace. (This account will be modified below, p. 318, but will suffice for our present purposes.) Now any entity that can be referred to at all can be referred to by some appropriate NP, and these entities certainly include propositions: the italicized NP in "*What you just said* surprises me" picks an entity out of the relevant universe that has the property of having just been said by the hearer, and which must presumably therefore be a proposition. I am suggesting that there is a rule to the effect that the universe of entities available to a given NP, out of which it picks one, never includes the proposition that is expressed by the sentence superordinate to that NP. Thus the subjects of sentences like *Everything I say is important* or *All propositions are either true or false* will be taken to refer respectively to all propositions expressed by the speaker other than the one expressed in the former sentence, and to all propositions other than the one expressed in the latter sentence — even though it would be in no way paradoxical if the assertions covered the latter propositions also. If *A* or *B* are uttered with silence preceding and following, then the only entity having the properties invoked by the wording of the subject NP is exactly that entity which is barred from the universe available to the NP, hence the NP, although having the form of a referring expression, fails to refer. I assume that the failure of an NP purporting to refer actually to do so is a sufficient condition for the utterance of the sentence containing that NP to fail to make a statement. (It will be clear that I agree with Strawson's (1950) objections to Russell's Theory of Descriptions.)

Notice that if what I say is true, then Stroll's approach to the consisten-

cy problem, apart from giving the answer we do not want to hear, is not as reasonable in itself as it seemed at first sight. Stroll argues that the reason *A*, preceded and followed by silence, does not occur is that there will never be any practical use for such an utterance. If Stroll wished to deal with the fact that the subject of *B* does not cover p_B, he would have to assign that fact some quite separate cause. I have offered a rule according to which a solitary utterance of *A* is meaningless and the subject of *B* cannot refer to a set including p_B. Other things being equal, an account that subsumes disparate facts under a single generalization must be preferred to one that deals with them separately.

An obvious objection to my rule is that I have already admitted that the logician, hearing *B* uttered alone, may understand it as asserting a proposition about itself. Indeed, if it were impossible for anyone in any circumstances to take the subject NP of *A* as referring to a statement made by uttering *A* then no question of paradox could ever have arisen. I would answer that it is certain that the forms of natural languages can be used as a sort of code for expressing formulae of any artificial logistic system; but that if no artificial conventions are adopted for interpretation of the forms of (say) English, there remains a unique interpretation for the forms of English (other than sense-bearing lexical items) that is 'given' rather than adopted by convention. I call the use of English with this particular interpretation 'natural' English, and when I speak of a 'normal hearer' I mean one interpreting what he hears as natural English.

A clear example of the distinction between artificial and natural English is that the logician can use the *if* of English as a 'code' for the connective of material implication in classical propositional calculus, and can recognize a sentence such as *If grass is red, two plus two equals five* as a true sentence with *if* used in this way; but any natural interpretation of the sentence would render it nonsensical or false, since material implication is not a connective existing in natural English — or, I expect, in the natural interpretation of any other language — and the connective represented by *if* is of another kind. Of course it is a matter of convention that the particular logical connective represented by *if* in natural English has the phonological shape 'high front vowel followed by voiceless labiodental fricative'; but I claim that it is 'given' rather than conventional that natural English has some way of representing that connective.

The question whether English is consistent is interesting only if 'English' is understood to mean natural English; there are no constraints on the artificial systems for which the forms of English can be used as a code. Since there exist artificial logistic systems in which self-reference is

possible, it follows that one can use the forms of English or another natural language in such a way that self-reference occurs; but this will only be possible by adopting artificial conventions about the interpretation of English in addition to, or instead of, the principles that operate in the natural use of English.

I do not pretend to be able to offer an empirical test for deciding whether a given utterance is a natural sentence-use or not; I am sure no such test is available. I find the idea that there is a distinction between natural and artificial sentence-uses intuitively appealing; I can only invite the reader to consider whether my account of language, which happens to embody such a distinction, seems as a whole enlightening or the reverse. The situation is a familiar one in linguistic theory: syntactic analysis depends on distinguishing well-formed, 'competent' utterances from utterances that have been distorted by 'performance' factors such as hesitation, imperfect memory, emotional disturbance, etc.; but there is no empirical criterion, only the judgment of intuition, to decide to which class a given utterance should be assigned. One of Chomsky's important arguments has been to the effect that this situation in linguistics is a very usual one for a science to be in.

If the distinction between artificial and natural use of English sentences is accepted, then we can explain the fact that the Liar paradox has troubled people. It is a social convention that we expect sentences addressed to us to make sense, and we frequently supply deficiencies in the sentences we hear in order to get them to do so, without necessarily being conscious of what we are doing (for instance, when a foreigner or a child is speaking, or when a native speaker's utterances are distorted by performance factors). So it is understandable that although the normal reaction to a solitary utterance of *A* will be a blank look, a logically-minded person may temporarily adopt an artificial convention to the effect that the universe of entities available to the subject NP of *A* includes any proposition that may be expressed by *A* — in which case *A* will after all be able to express a proposition, and the paradox will arise, but not in 'natural' language.

If it is true that self-reference does not occur in natural language, then we have achieved our aim of dissolving the problem of the Liar paradox empirically rather than *a priori*. However, the situation can still not be regarded as completely satisfactory. That self-reference is impossible in natural language, if true, is highly convenient from the logician's point of view. But why should it be true, since there do not appear to be any other cases of limitations on the entities that can be referred to by appropriate

NPs? Can the prohibition of self-reference be derived from independently-known facts about natural language?

I believe it can. In the remainder of this paper, I first summarize some recent work, and then go on to show that the prohibition of self-reference can be deduced without further assumptions.

In recent works (Sampson, 1969a, 1969b) I have offered a view of the process of sentence-comprehension by a hearer according to which a heard sentence effects a change of state of an abstract automaton (representing an aspect of the hearer's mind), such that the particular new state reached depends on the deep structure of the sentence and on the previous state of the automaton. The nature of the automaton can be specified rather precisely; I shall here go into only such detail as is necessary for our present purposes.

In the first place, it seems clear that we are forced to admit that the relation of 'reference' between NPs and entities in the world at large is a composite relation. Rather than two sets of relation-terminals, NPs and entities-in-the-world, there must be three sets, namely those two and an intermediate set of abstract entities in a hearer's mind, entities such as Geach's "Ideas" (Geach, 1957: 53ff). One relation, which I shall call 'reference$_1$', has NPs as domain and hearer's-mental-entities as range; while a separate relation, 'reference$_2$', has the latter as domain and entities-in-the-world as range. The relation commonly called 'reference' is the composition of reference$_1$ with reference$_2$. The need to admit the intermediate set of hearer's-mental-entities has already been argued within linguistics by Karttunen (1968); and cf. McCawley (1968). If it were the case that the relation between NPs and outside-world objects were a simplex relation, it would appear impossible for works of fiction to exist. Instead of Dostoevsky having written that Raskolnikov killed Alëna Ivanovna, he might equally well have written that Alëna Ivanovna killed Raskolnikov, since both the NPs *Raskolnikov* and *Alëna Ivanovna* would refer to the null set. But we know that it makes a great deal of difference that Dostoevsky wrote one of these rather than the other, since, although the NPs in question have identical (null) reference with respect to the outside world, they have differing reference so far as the reader is concerned. So they must refer to entities within the reader's mind. (A similar argument could be given with respect to NPs denoting abstractions like *the metric system* or *what John said*, where the question whether any referent for the NPs exists in the world seems scarcely meaningful.)

Of course, it may be objected that it is possible to set up a logistic

system in such a way that the formal translation of, e.g., *Raskolnikov killed Alëna Ivanovna* does not in fact imply that *Raskolnikov* has reference. Thus Reichenbach devotes a section (1947: 49) to the problem of the formal representation of fictitious existence, without postulating a level of mental entities. But note that such an avenue of escape is available only to one who, like Reichenbach, is establishing his semantic representations for natural-language sentences solely by reference to introspective judgments of meaning. We, on the other hand, are obtaining our semantic representations from empirical syntactic research. In order to maintain that natural-language deep structures can be interpreted without invoking a level of mental entities, it would be necessary to claim that syntactic evidence can be found for the hypothesis that the deep structure of *Raskolnikov killed Alëna Ivanovna* differs from that of, say, *Oswald killed Kennedy*. It is true that in many cases syntactic evidence can be found to show that constructions that superficially look alike in fact derive from different underlying structures; but I cannot think of a shred of syntactic evidence that would support differing derivations in the *Raskolnikov* and *Oswald* cases. The onus is on whoever wishes to argue against mental entities to provide such evidence.

I shall therefore take it for granted that one property of the automaton representing the hearer's (or reader's) mind is that it contains some kind of array of entities (I call them 'referents'), which can be referred to by NPs. Clearly the automaton must store information in some form about the properties of each referent, since NPs often succeed in referring by means of describing the entity in question *(the short, fat man)*. If a mental referent *a* corresponds to a physical object that has the properties of being short, fat, and a man, I shall say that *a* itself has the properties §(short), §(fat), and §(man). (Here I adopt Geach's [1957: 52f] operator '§( )' to create properties for mental referents corresponding to properties of entities in the world at large. A mental abstraction cannot be fat, but it can be §(fat).) Clearly many of the referents will correspond to entities that do exist in the outside world, and these referents will have the property §(existent) or §(real).

Thus if I use the expression "the handkerchief" to you and you understand me to be indicating a particular handkerchief, the relation between the expression "the handkerchief" and the handkerchief is a composite one, comprising a relation of reference$_1$ between the NP and some one of your mental referents that has the property §(handkerchief), composed with a relation of reference$_2$ between that referent and the physical object. Other expressions, such as "Raskolnikov", will correlate with nothing in

the world at large; but only the reference$_1$ relation is necessary for communication to occur.

Now I would argue that referring NPs must be divided into two classes, which I call 'identifying NPs' (INPs) and 'establishing NPs' (ENPs). The INPs include proper names, pronouns, and NPs that begin with *the*, a deictic (*this*, *that*), or a possessive (*my*, *John's*). INPs have the function of picking out a particular referent among the array of referents already present in the hearer's automaton. Thus the sentence

E. John bought the red car

will make sense (in one sense of the phrase 'make sense') only to a hearer who already knows of the existence of a person called "John" and of a red car. Earlier, I suggested that for *E* to make sense there must be only a single referent describable as a "red" "car" (i.e., only one referent with the properties §(red) and §(car)) in the universe of referents available to the NP. In fact, of course, *E* may make sense to a hearer who is aware of several red cars, as long as one of them is somehow nearer to his focus of attention than the others. So we must suppose that the hearer's referents are arranged in some kind of abstract space, with one point in that space singled out as the 'focus'. Simplifying a great deal, we may then say that an INP picks out a single one of the hearer's referents according to the following rules: if the INP is a pronoun, it picks out the nearest referent to the focus; if it is a proper name, it picks out the nearest referent to the focus bearing that name (more precisely, the nearest referent bearing the reference$_2$ relation to an entity bearing the name); if it is a phrase beginning with *the*, it picks out the nearest referent to the focus having the properties specified by the lexical items following *the*. Thus the reason why, in the sentence *D* (p. 313), the second act of refer-ring$_1$ requires only the pronoun *it* while the first uses a number of lexical items is not that the universe of referents has shrunk between the utterance of the two INPs, but because the focus has come closer to the referent that refers$_2$ to Buckingham Palace. If the focus represents the center of the hearer's attention, it is intuitively reasonable to suppose that refer-ring$_1$ to a referent early in the sentence will bring the focus close to that referent for the rest of the sentence. Obviously, in a full discussion it would be necessary to specify formally the nature of the referent-space and the measurement of distance within it, and the precise way in which the position of the focus is determined; for these matters I must refer the reader to Sampson, 1969b.

Some phrases having the form of INPs will fail to pick out any referent,

of course; thus *the hungry octopus* said to a hearer who has no knowledge of any hungry octopuses either in reality or fable, or *the present king of France* said to anyone in 1970, will simply fail to refer$_1$. To utter a sentence containing such an NP to such a hearer is to fail to make a statement.

Notice particularly that referents can correspond not only to atomic entities but also to propositions; but an INP referring to a propositional referent often starts with *that* rather than *the*. Thus the object of the sentence

F. Roger discovered that John bought the red car.

is an INP. The referent, call it *a*, to which that INP refers$_1$, has the property §(act of buying) — or we may simply write '§(buy)' — and dominates an ordered pair $<b, c>$, where *b* is the referent identified by *John* and *c* that identified by *the red car*. (I neglect the matter of tense.) The referent *a* is called a 'propositional referent', while *b* and *c* are 'atomic referents'. The term 'dominate' is a primitive term of the theory: intuitively, the referents dominated by a propositional referent are the entities that proposition is about (whether atomic referents or further propositions). A propositional referent may dominate various numbers of referents, e.g., the referent picked out by the INP *that John is happy* will dominate just the referent *b*, whereas the referent picked out by *that John sold Mary a brooch for £1* presumably dominates an ordered quadruple of referents.

Notice that if one had already been discussing John's buying the red car in previous sentences, and hence *a* was close to the focus, then instead of saying *F* one could communicate the same thing by saying *Roger discovered the fact* or *Roger discovered it*, where the lexical specification of the target referent *a* is very meagre.

The properties of automaton-referents that may be used by *the*-INPs to pick out their target referent include not only the simplex properties indicated by individual lexical items (nouns, adjectives, verbs) but also the domination-relations into which the referent enters, which are specified by relative clauses. Thus if the hearer has been told the sentence *E, John bought the red car*, then his referent *c* can be identified by means of the INP *the car that John bought*. This INP invokes the fact that *c* has the simplex property §(car), and also that *c* occurs in an ordered pair with *b*, and that this pair is dominated by a referent having the simplex property §(buy).

Whereas an INP picks out one of the referents in the hearer's automaton, an ENP creates a new referent in the automaton. The NP *a red car* in

G. John bought a red car.

is an ENP; one who hears *G* addressed to him will add a referent, say *d*, to his automaton, and store the information that *d* has the properties §(red) and §(car). Furthermore, the declarative sentence itself is an ENP. Thus the utterance of *G* will involve not only the identification of *b* (*John*) and the creation of *d* (*a red car*) but also the creation of a propositional referent, say *e*, with the property §(buy) and dominating the ordered pair $<b, d>$.

When a sentence is complex, in the sense of having subordinate clauses embedded in superordinate clauses — e.g., the sentence

H. The intelligent teacher knew that Bruce disliked the problem.

which has the structure [*the intelligent teacher know* [*Bruce dislike the problem*]] (again I neglect tense for simplicity) — the action of the sentence on the hearer's automaton will begin with the most subordinate clause and work upwards. Thus the effect of *H* will be as follows: *Bruce* picks out a referent *f*; *the problem* picks out the nearest §(problem) referent, *g*; then a referent *h* is created, with the property §(dislike) and dominating the pair $<f, g>$; the nearest referent *j* having both properties §(intelligent) and §(teacher) is picked out; and finally a referent *k* is created having the property §(know) and dominating $<j, h>$. (It appears that clauses beginning with *that* may be ENPs or INPs without alteration in form. Thus I have assumed that the clause *that John bought the red car* in *F* is an INP, and that *a* refers$_2$ to a fact already known to the hearer, whereas *that Bruce disliked the problem* in *H* is an ENP, so that the hearer learns the fact to which *h* refers$_2$ by hearing *H*; but if the hearer's automaton does not in fact contain *a* then the *that*-clause of *F* will establish it, and if the hearer's automaton does contain *h* the *that*-clause of *H* will identify it.)

A final point to be made is that, although I have been discussing sentences that at the surface are simple clauses, e.g., *Mary washed the plate*, as if they were simple clauses also in deep structure, in one important respect this is believed not to be so. Ross (1970) and others have given evidence that at the deep level all sentences have a superordinate clause containing a performative verb, though this clause will frequently be deleted transformationally. Thus the deep structure of *Mary washed the plate* will be similar to that of *I assert that Mary washed the plate*, the deep structure of *Go home!* will be similar to that of *I order you to go home*, and so on. So a full account of the change of automaton state of one hearing *Mary washed the plate* will be as follows: *Mary* identifies a

referent *m*; *the plate* identifies a referent *n*; a referent *p* is created with the property §(wash) and dominating the ordered pair $<m, n>$; *I* identifies a referent *q*; and a referent *r* is created with the property §(assert) — or, equivalently, §(say) — and dominating the ordered pair $<q, p>$.

This account is simplified to the point of distorting some of the facts about language; and even the fuller version in Sampson, 1969b can be described only as a plausible hypothesis, rather than as an established theory. But the theory of language offered there, for what it is worth, was composed without considering questions of self-reference or paradoxicality at all.

It turns out as an unexpected bonus of the theory that it accounts for the lack of self-reference in natural language. Consider a sentence purporting to be self-referential: e.g. sentence *B*, *What I am now saying is not to be repeated.* The deep structure of this sentence would be along the lines of [*I say* [[*that which I am now saying*] *is not to be repeated*]]. What sort of automaton change-of-state should such a deep structure correspond to? To start at the end, clearly the superordinate clause must create a new referent, say *s*, with the property §(say) and dominating an ordered pair of lower-level referents. One of these will be the referent identified by *I*, say *t* — there is no problem here. The other will be a propositional referent created by the clause *that which I am now saying is not to be repeated.* Let us call this referent *u*. Note that *u* must exist in the hearer's automaton BEFORE *s* can be created, because *s* must dominate $<t, u>$. Now *u* will be a referent with the property §(not to be repeated) and dominating the referent identified by the phrase *that which I am now saying*; and hence this latter referent must be identified before *u* can be created. How will the phrase *that which I am now saying* pick out a referent? It should pick out the nearest referent *x* to the hearer's focus having the property specified by the relative clause, i.e., such that some referent *y* with the property §(is now saying) dominates $<t, x>$. Can *x* and *y* be identical to *u* and *s* respectively? If *u* and *s* already existed, they could — when *u* and *s* have come to exist in the hearer's automaton they will have the properties required for *x* and *y* (assuming the properties §(say) and §(is now saying) are equivalent). But at the time the NP in question is trying to find a candidate for *x*, the referent to be created by the clause including that NP (namely *u*) cannot yet exist (and so *a fortiori s* cannot exist either). If the speaker has just uttered some OTHER proposition or set of propositions *w* sufficiently recently for the tense of *am now saying* to cover the utterance of *w*, then *x* will be identified with *w*, the new referent *u* will dominate *w*, and *B* will succeed in creating a referent *s* dominating

$<t, u>$. This is what I suggest happens in the normal, parenthetical use of *B*. But if *B* is uttered without accompanying sentences, the phrase *what I am now saying* will simply fail to refer. It must pick out an existing referent before the referent *u* can be created; yet other than *u* there will be no candidates with the specified property. Since the phrase *what I am now saying* will fail to refer, *u*, which needs a referent to dominate, can never be created; and hence nor can *s* in its turn. The sentence *B* will fail in its task of creating a propositional referent in the hearer's automaton; and so will any other sentence in which an embedded clause purports to refer to a referent awaiting creation by a superordinate clause, including *A*.

So the problem of the truth of *A* entailing the falsity of *A* and *vice versa* will never arise. *A* will be neither true nor false, just as *The present king of France is bald* is neither true nor false. The possessors of truth-values are not utterances, but propositional referents created in automata by utterances. *A*, and *The present king of France is bald*, fail to create referents, so that no entities exist to have truth-values. There will, admittedly, be a difference between these two sentences. *The present king of France* might refer if spoken next year, if France happens to become a monarchy: the lack of reference of this phrase is a merely contingent fact. On the other hand it is impossible that the phrase *what I am now saying* could ever refer, as long as the sentence within which it is embedded is uttered alone (unless one or more of the vocabulary items it contains, e.g., *now*, *say*, comes to change its sense).

I conclude that, whatever possibilities of inconsistency may occur in natural language, the paradox of the Liar is not one of them.

REFERENCES

Anscombe, G.E.M.
1963 "Review of Ziff, *Semantic Analysis*", *Mind* 72, 288-93.
Bar-Hillel, Yehoshua
1966 "Do Natural Languages Contain Paradoxes?", *Studium Generale* 7, 391-97.
Chomsky, Noam
1961 "On the Notion 'Rule of Grammar'", *Structure of Language and its Mathematical Aspects* (= *Proceedings of Symposia in Applied Mathematics* 12), ed. by R. Jakobson (Providence: American Mathematical Society), 6-24.
1965 *Aspects of the Theory of Syntax* (Cambridge, Mass.: MIT Press).
Geach, P. T.
1957 *Mental Acts* (London: Routledge and Kegan Paul).
Karttunen, Lauri
1968 "What Do Referential Indices Refer To?" (= *RAND Corp. research paper* P-3854) (Santa Monica).

Katz, J. J.
1966 *The Philosophy of Language* (New York: Harper and Row).
1967 "Recent Issues in Semantic Theory", *Foundations of Language* 3, 124-94.
Katz, J.J. and J.A. Fodor
1963 "The Structure of a Semantic Theory", *Language* 39, 170-210.
Lakoff, George
1970 "On Generative Semantics", to appear in *Semantics*, ed. by D. D. Steinberg and L. A. Jakobovits (London: Cambridge University Press).
McCawley, J. D.
1968 "The Role of Semantics in a Grammar", *Universals in Linguistic Theory*, ed. by E. Bach and R. T. Harms (New York: Holt, Rinehart and Winston), 125-69.
Postal, P. M.
1970 "On the Surface Verb *remind*", *Linguistic Inquiry* 1, 37-120.
Quine, W. van O.
1953 *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press).
Reichenbach, Hans
1947 *Elements of Symbolic Logic* (New York: Free Press).
Ross, J. R.
1970 "On Declarative Sentences", *Readings in English Transformational Grammar*, ed. by R. A. Jacobs and P. S. Rosenbaum (Boston: Blaisdell-Ginn), 222-72.
Sampson, G. R.
1969a "Noun-phrase Indexing, Pronouns, and the 'Definite Article'", *Yale Linguistic Automation Project Report* (New Haven).
1969b "Towards a Linguistic Theory of Reference", PhD dissertation (Yale University).
1970 "Review of D. T. Langendoen, *The Study of Syntax*", *Journal of Linguistics* 6, 267-77.
Forthcoming "The Concept 'Semantic Representation'", Paper read to LAGB Autumn Meeting, University of East Anglia, 1 November 1970.
Strawson, P. F.
1950 "On Referring", *Mind* 59, 320-44.
Stroll, Avrum
1954 "Is Everyday Language Inconsistent?", *Mind* 63, 219-25.
Ziff, Paul
1960 *Semantic Analysis* (Ithaca: Cornell University Press).

Geoffrey R. Sampson (b. 1944) is a Fellow of Queen's College, Oxford. His principal research interest is theoretical linguistics, and his most recent publication is "On the Need for a Phonological Base", *Language* 46 (1970), 586-626.