J. P. THORNE

information; its manifestation, however, is not an intellectual failure but loss of mastery of a perceptual technique.[1]

A grammar will provide the most rigorous imaginable series of tests for a performance model; it will specify the results a performance model must give in every case. But the assumption underlying much recent work in psycholinguistics, that the study of performance models, is the right place to begin the study of performance models, is surely misleading. As long as this assumption is made we are in danger of asking questions to which there are no answers.

Discussion

JOHN MORTON

In discussing Thorne's paper, I find myself at a severe disadvantage. I do not, for example, unlike Wittgenstein, find it queer that there should be such a thing as 'hearing a word in a particular sense' (only that there should be such a sentence as *Hearing a word in a particular sense*), nor do I find it remarkable that our use of the word *hear* is systematically ambiguous (any more than I am surprised to discover that the word *table* is systematically ambiguous). Indeed I would claim that *hear* is no more ambiguous than the average everyday word. It is only when we try to use the word in a technical context that we may be troubled. I cannot tell what Thorne understands by 'the ordinary sense' of the word *hear*. Is it a descriptive word referring to the situation of a person whose ears and auditory cortex are being stimulated or does it refer to some internally-recognized, introspectible event which leads a person to say *I hear*.... If the latter, then we have no need to specify what operations have been performed upon the input in order to use the word, since the nature of the operation will be apparent (in principle) from the continuing phrase. Thus *I heard a sentence* implies that an internal event has occurred which was related (or was supposed by the system to be related) to auditory stimulation and concerning which I have responses available. *I heard a noise* implies exactly the same. In the former case the verbal responses (there may be other relevant responses available but these will not be considered) include a string of words, which a third person would consider to be the 'same' as the original string, and des-

1. It is interesting to speculate on the results of performing Fodor and Bever's experiment with aphasics—also with young children. One would guess that in each case the movement of the click would follow a quite different pattern from that observed in the case of normal adults.

criptions of various 'acoustic' parameters such as intonation, stress, voice, quality and the like. The experiment by Lieberman (1965) to which Thorne refers shows that, in some circumstances, such descriptions bear only a vague relationship to physical measures of the stimulus. When we say *I heard a noise* our description of the stimulus is likely to have a closer relation to the actual structure (with qualifications concerning the amount of information to be processed), but, none-the-less, the acoustic stimulus must have undergone complex processing for us to be able to say it was a *noise* and not a *train* or a *bird*. *Noise* is a label as much as *train*. I agree that *hearing a sentence* involves, if you like, monitoring one's own response, but this is incidental to the use of *hear*. St Joan *heard* her voices—the system made an error, though at a higher level.

I do not intend what I have said to be taken as a philosophical criticism of Thorne's paper. I have no philosophical training and, although I appreciate the complexity of the philosophical problems he raises, I cannot pretend to appreciate the problems themselves. My first point is rather that such problems are not relevant to the work of psycholinguists, or, rather, have not been shown to be relevant; and I would be surprised if a discussion of the logical inconsistencies of natural languages furthered our ends.

These ends, I take it, are to understand the mechanisms involved in language behaviour; and the method we use is to attempt to simulate this behaviour with models (of various degrees of specificity), computers, equations, or rules. If what Thorne is saying is that we cannot usefully talk about these mechanisms in natural language, then I agree. There are many psychologists who talk about 'perception' as if it were a single operation that produced some THING which was then reacted to—as if our behaviour was a series of 'double takes'. This leads to such statements as 'The word-frequency effect is all a response effect', or 'Perceptual defense is all in the stimulus', which, not surprisingly, are refuted in the following year by someone who holds the opposite viewpoint. This usage may be akin to Thorne's use of *hear* and *interpret*. However, I would criticize such theories by attempting to point out that the underlying mechanisms could not work. Thorne appears to criticize analysis-by-synthesis mechanisms by reference to linguistic anomaly. He described analysis-by-synthesis as an approach 'which treats UNDERSTANDING as the application to what is HEARD of an INTERPRETATION formed from a knowledge of linguistic structure (p. 7: my emphasis). But the words *understanding*, *interpretation* and *heard* are his own and do not describe the devices suggested either by Halle & Stevens

(1964) or by Matthews (1961). It seems paradoxical to point to Halle & Stevens' flow diagram and say that a particular part of it is ('corresponds to', implies) 'hearing', or to say of Matthews' proposals that they constitute 'interpretation', or of either that their aim is 'understanding', and then criticize the models on the basis of these WORDS.

While I agree with Thorne's conclusion that analysis-by-synthesis is unlikely to provide us with a satisfactory performance model, it seems to me that the only valid way of criticizing such models, and the principles behind them, is in terms of the mechanisms: what they cannot do and whether they are necessary. This turns out to be a semantic trap in the end; for although analysis-by-synthesis is initially defined as matching some internally-derived signal or symbol string with an input string, both Halle & Stevens and Matthews (unlike the Haskins group, with their 'motor' theory of perception: cf. Liberman, 1957; Liberman, Cooper, Harris & McNeilage, 1962) so modify their proposals in terms of 'preliminary analysis' that a strict 'matching procedure' becomes a very small part of the operation. Indeed, Halle & Stevens (1964) go so far as to say: 'A rough preliminary analysis at each of the stages...may often be all that is required—ambiguities as a result of imprecise analysis at these early stages can be resolved in later stages on the basis of the constraints at the morphological, syntactical, and semantic levels'.

Doubts as to the amount of synthesis in the system arise from considering how there can be a rough preliminary analysis in a system which produces an 'Output Phoneme Sequence' and which utilizes 'generative rules for transforming phoneme sequences into phonetic parameters'. Is the Output Phoneme Sequence likely to be incorrect? Is it likely to give positive but misleading information? It seems much more likely that it will give information of the form 'this phoneme is either /p/ or /b/ or /m/'. Such a statement could be made more economically by specifying, for example, the confirmed distinctive features, leaving the others unspecified. In this case we do not need 'trial phoneme sequences' from which to generate a match, indeed they would be inappropriate; for to produce a statement '/p/, /b/ or /m/' requires that the common features be confirmed in isolation. Thus their strategy system either has to produce for comparison all phonemes in all positions (if the analysis is imprecise, constraints will not be of much use at this stage) and select those with the best (preliminary) fit; or it merely has one attempt and passes information concerning the errors computed in the comparator on to the next stage. This error is supposed to control which phoneme sequence, or set of phonetic parameters, is next presented for comparison. It must therefore contain information about which parameters are

NOT satisfied. If the system can do this then we might expect it to be able to test for the presence of each feature (or other element) directly in a manner similar to a picture grammar (cf. Clowes, 1966). Do we make the kinds of recognition error which could be expected from such a system? Is this the nature of the 'rough preliminary analysis', and can this really be called 'analysis-by-synthesis'?

Lest it be thought that I am quibbling over terms, let us consider a later stage in the analysis—we will assume that the morphemes have been recognized (at least partially) and that the surface and deep structure analysis systems are ready to go into action (cf. Matthews, 1961; Thorne, Dewar, Whitfield & Bratley, 1965; Wales & Marshall, in the present volume). Thorne rightly states the basic postulate of analysis-by-synthesis: that 'understanding an utterance results from the hearer constructing a sentence to match it.' But is it any more necessary to match the sentence than it is to match a phoneme sequence which is a part of that sentence? If we consider the existence of constraint at the latter level, we must also consider it at the former level; and if we can show that certain elements of a sentence, without regard to syntactic structure, are sufficient to make it unambiguous, then why go to the trouble of fully matching it? Thus if I say, *Cat sat on mat—Fire spat—Fur flames—Cat withdrew—Quick—Ouch*, no one has any trouble in understanding what I am saying. Equally we can consider the example of the reversible passive given by Wales & Marshall (p. 71), of which, however, I offer a different account. Wales & Marshall suggest that since *The boy...hit...the girl* is semantically well formed, the implications of *was* and *by* are not explored. *The engine...repaired...the mechanic*, on the other hand, being semantically deviant requires, and so receives, a complete syntactic analysis. I would rather say that the string *engine, repaired, mechanic*, being non-ambiguous, will be acceptable to the suprasemantic system without further analysis, the LOGICAL structure of the original sentence (active or passive) being implicit in the string. Given *boy—hit—girl* on the other hand, we need further information in order to obtain a unique result, and so we take longer to process the sentence. Both accounts are compatible with Slobin's (1963) result. We might also note that *house build young* is not in this sense ambiguous, whereas *man—house—build—large* is. So we might expect to find differences of re-action to sentences of the form:

The young man built the house

and

The large man built the house.

The understanding of the latter sentence requires some registration of the word order (since the adjective could apply to either noun), whereas the understanding of the former does not.

Does the system become an 'Analysis-by-Synthesis System' merely because it can utilize analysis-by-synthesis to iron out residual ambiguity, and is it necessary to think of the system in any other way? Katz & Postal (1964: 168) go a stage further and say that in the light of their analysis 'the understanding of a sentence... is obtaining its semantic description' (my emphasis). If they wish merely to define the word *understand* in this way, then there is nothing more to be said, but such a definition is not useful. If they are using the word in its more usual (albeit imprecise) sense, then it is clear that 'understanding' is more than this, and 'understanding' is not necessarily all of this; and I am talking here in terms, not of the logic of the system, but in terms of the mechanisms (rules, processes) which operate. Thus when Thorne says 'to understand an utterance is IN SOME SENSE to know its structure' (my emphasis), this is only acceptable as a logical statement. It need not, and should not, be interpreted as saying 'to understand an utterance we always have to apply all the rules we would use in producing the sentence'. Given *house—build—man*, we do, in some sense, know the structure of the underlying sentence structure and can produce it.

If we wish to say that 'understanding' is something, then it is necessary to go beyond the semantics of the immediate sentence and include some conceptual system which deals with objects, events, actions, and so on.

Compare:
1. 'The understanding of a sentence...is obtaining its semantic description.'
2. '"Pol...achi...lumn", he gasped with his last breath, but this was enough for Nigel to know that the Polish waitress who had served him that day with borscht and vodka was in real life the maiden aunt of the engineer who had invented the machine which would revolutionize potato farming and that his mysterious bill for 78 groats was in fact a secret message instructing him to meet them at the base of Nelson's Column in order to foil the plot of their natural enemy.' That is understanding!

I wish to emphasize that I am not, at the moment, trying to claim that the brain never performs analysis-by-synthesis, merely that it is not necessary to assume it always, or even usually, does. I would accept the possibility of analysis-by-synthesis at the phonetic level under certain conditions (Morton & Broadbent, 1964), and I can see that there may

14

be good reasons for including a synthesis system in a syntax recognizer, though the algorithm for surface structure analysis developed by Thorne, Dewar, Whitfield & Bratley (1965) constitutes a very strong counter-argument.

Having suggested that in hearing sentences we do not necessarily use all the relevant structural rules, I would like to make what may turn out to be a similar suggestion for our production of sentences. In producing a sentence we do not usually start with S, and proceed to derive NP and VP, etc. Instead, some event occurs in what Wales & Marshall call the 'Conceptual Matrix'. Such an event may be of the form of a desire to express a relationship between *house*, *man*, and *build* (of course we need to develop a precise descriptive terminology for such events). We could express this relationship in either the active or the passive form. Since there is no received performance model of production, I am uncertain as to how deviant my viewpoint is; but part of what I think may happen is that whichever of *house* or *man* is first available as a response (in some speech motor code; cf. Morton, 1964a, b), this becomes the surface subject, and the decision as to whether the active or the passive form is used is determined by feedback from the phonological system to the Surface Structure Deriver (cf. Wales & Marshall, in the present volume: p. 55). My reason for suggesting this scheme is that we seem, where possible, to make the surface subject a word which has previously occurred. Thus we would say:

I saw the house. The house was built by the man.

or

I saw the man. The man built the house.

but not:

I saw the house. The man built the house.

or

I saw the man. The house was built by the man.

The situation seems analogous to the observation that when a word has been recognized or produced as a response, it is more likely to be available as a response in the near future (Morton, 1964c).

Finally, I would like to try and link together two problems which Thorne raises: ambiguity and the relation between sound and structure. Thorne examines what he calls 'the ambiguous utterance': *I dislike playing cards*. When I started to say this sentence, however, I discovered that

15

I could not say it ambiguously; that, when spoken rather than written, it did not seem ambiguous. I invited a few of my colleagues to try and utter the sentence ambiguously, and then had them judge each other's efforts. The result was that all the observers agreed in their judgments of the sentence. Now, assuming that this little experiment can be repeated, the result should not surprise anyone. If we use structural rules when we produce a sentence, we must use one or other of the two alternative sets of rules in uttering the sentence *I dislike playing cards*. If then certain aspects of structure are, under certain conditions, reflected in the acoustic pattern of the semantic utterance, the result found would follow. I do not mean to say that the structure of an utterance is labelled by a series of physical cues any more than I want to say that the phonological structure of an utterance is labelled by a series of physical cues. The two are merely correlated. Thus Lieberman (1965) showed that when a redundant word was gated out of its containing sentence and was presented in isolation, it was less intelligible than the same word gated out of a non-redundant context. Equally, in a one-subject experiment, people asked to make the *transport—transport* distinction by saying the word in isolation distinguished between the two very clearly by means of pitch, intensity and duration. When the words were embedded in a sentence, however, the distinction was not apparent in the isolated word.

These observations lead to the proposition that ambiguity, of certain kinds, only occurs within a section of an utterance when the rest of the utterance is sufficient to make that section unambiguous. This seems to imply the same order of feedback in the speech-production system as we have in the recognition system. For Thorne's other examples, these considerations do not seem to apply; and I share his interest and delight in ambiguity. I would like to present two further examples for your consideration. In the course of an experiment, I required subjects to complete sentences, one of which was *He asked the way to the....* I spoke the sentences conversationally, and typical completions for this example were *station, hospital, bus-stop.* One subject replied, and wrote, *time.* Even saying this apparently schizophrenic sentence aloud, most people take some time to make sense out of it, despite (or because of?) the fact that little or no phonological distortion is required. In contrast to this example, the four words *Whale oil bead hammed* are readily restructured, in spite of the fact that the change involves adopting a strong Irish accent.

To conclude, I agree with Thorne that structure must be involved in hearing (though not necessarily in 'hearing'); I agree with his assessment of the dangers of equating competence with performance models;

I agree with him that analysis-by-synthesis will not provide an adequate account of speech recognition (although not, on the surface, for the same reason); but I do not really understand the nature of many of his questions, nor what kinds of answers he requires. If I have answered any of his questions accidentally, I will be very happy; but in any event I hope I have been able to make clear the questions which interest me.

R. C. OLDFIELD

I must start by confessing that I find myself in something of a difficulty in commenting on Thorne's admirable and stimulating paper. To me, the ensemble of words like *hear, understand, recognise, interpret* presents so baffling and dimly appreciated a semantic web that any attempt to clear my mind about it would take all day, and would probably leave us all more confused than before. I do not suppose that the minds of other people here, in particular, Thorne's, are nearly as muddled as this, because in the course of discussion they will have sharpened the senses in which they are using these words and, at any rate implicitly or intuitively, will have got a grip on them. I would like first to give only one illustration of the kind of difficulty I find myself in before going on to offer a few comments on Thorne's views so far as I understand them.

Consider the word *interpret*. Thorne makes use of this word in confronting two views as to what happens when two identical physical stimuli provoke, in two different organisms, different internal states or responses: 'this [Lieberman's] finding runs quite contrary to the idea that understanding an utterance can be described as a two-stage process of INTERPRETING what we hear' (p. 5; my emphasis). When I try to think what I understand by the word *interpret*, what do I find? First, I turn to the usage suggested by the image of 'an interpreter', as it might be at a conference between two statesmen with no common tongue, or in a court of law when the defendant is a foreigner. This does not seem to help very much as his function is expressly that of recording as accurately as possible what has already been said in one language into another. He is a mere go-between or 'transducer'. Secondly, I think of the expression 'reading between the lines'. Here we have a situation in which a given stimulus already evokes, in accordance with existing experience or convention, one meaning which is, however, rejected in favour of another in consequence of the recipient being shrewder, more sceptical by nature, or taking into account particular features of the situation or particular special knowledge he possesses. Thus, I might say: 'He is going to the Paris Conference after all', which I INTERPRET to mean (or which, BEING INTERPRETED, means) 'we were right about