PSYCHOLINGUISTICS

JOHN MORTON Ph.D.

Medical Research Council
Applied Psychology Unit
Cambridge

It used to be believed quite generally that our use of language could be studied adequately merely by observing verbal behaviour. This belief reached its apogee, perhaps, when Skinner (1957) wrote:

> The basic processes and relations which give verbal behavior its special characteristics are now fairly well understood. Much of the experimental work responsible for this advance has been carried out on other species, but the results have proved to be surprisingly free of species restrictions. Recent work has shown that the methods can be extended to human behavior without serious modification.

This statement appeared to many to be untrue at the time and it remains so. We still have only the vaguest understanding of the basic processes of language, and there is little or no experimental work with animals with the slightest relevance to the problem.

In the 1950's, many psychologists who did not believe Skinner thought, also mistakenly, that the techniques of information theory were the key to understanding language. However, neither Skinner's position nor information theory has been of much assistance since then, except in demonstrating the limitations of the respective approaches. The year 1957 also saw the appearance of Chomsky's *Syntactic structures*, in which he sketched the earliest version of his theory of grammar. This and subsequent versions of the grammar gave the promise of providing a new model for language processing. The excitement which followed has somewhat diminished, partly because of the failure of attempts to apply Chomsky's grammar directly to behaviour. However, there has been an increasing realization of the complex and indirect relation between our behaviour and anything we might come to call the language stimulus. We will trace these relations at several stages of the language process.

1.   Speech Perception

The shortest significant segment of speech is called a "phoneme". Phonemes are not defined absolutely but in relation to each other, the three phonemes in "cat" ([k], [æ] and [t]) being established by contrasts with (for example) "mat", "cot" and "cad" respectively. These distinctions do not just depend upon the extent of differences between sounds. For example the [p] sounds in "pit" and "spit" are widely different, that in "pit" being aspirated. In some languages (e.g. Thai) an initial [p] can be aspirated or not—the difference leading to different words. This distinction is thus phonemic in Thai. Since they are both the same phoneme in English it is difficult for native English speakers to tell the difference.

The phoneme corresponds roughly to a letter but, whereas in the printed word one can point precisely to the boundary between one letter and the next, it is not possible to demarcate the boundaries between phonemes. Equally, the form of a printed letter is not dependent upon the surrounding letters, but the acoustic form of a phoneme is greatly influenced by the nature of the surrounding phonemes. Figure 1a shows spectrograms of "dee" and "doo", said by me. The amount of acoustic energy at a particular frequency and time is indicated by the darkness of the lines. Figure 1b shows those parts of the sound which are essential for the recognition of the syllables (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). The lines indicate the positions of the first two formants —frequency regions in which acoustic energy is concentrated. The right-hand part of each pattern would be sufficient to specify the respective steady vowels. At the beginnings of the patterns the centre frequencies of the formants are changing. The change in the lower, or first, formant serves to distinguish [d], together with [b] and [g], from the voiceless

FIG. I.   The sounds "dee" and "doo"
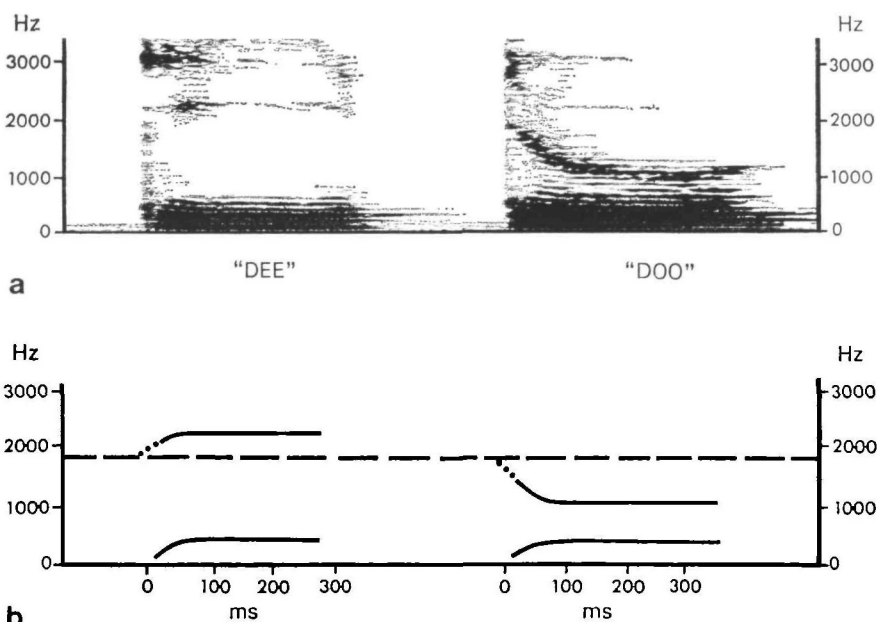
a: spectrograms of the sounds "dee" and "doo". The dark areas show presence of acoustic energy at the frequency indicated by the ordinates

b: idealized representation of the sounds, showing the parts necessary for perception. The interrupted line at the 1800-Hz level represents the resonance frequency of the vocal cavity at the instant of closure. For an explanation of the dotted portions of the curves, see text (section I)

195

[*p*], [*t*] and [*k*]. It is the changes in the second formant which serve to differentiate [*d*] from [*b*] and [*g*]. Although the [*d*] is perceptually the same in the two cases, the acoustic cues which serve the percept are widely different. The invariant feature of syllables with an initial [*d*] appears to be that the second formant "points" at, but does not actually reach, a common locus of 1800 Hz, as indicated by the dotted lines in fig. 1b. The reason for this lies in the way the sounds are articulated. In all syllables beginning with [*d*] the articulatory tract is closed at much the same place. It has been estimated that the resonance frequency of the vocal cavity at the instant of closure is about 1800 Hz. However, since no sound emerges until a certain amount of opening has occurred, the locus associated with [*d*] remains imaginary (in terms of sound) and thus would have to be calculated by the speech-recognition system rather than directly observed. The phenomenon—of hearing the same initial sound—is thus somewhat removed from the stimulus. If the second formants of the two sounds are heard without the first formant, they sound completely different from each other and totally non-speech-like. Thus the interpretation of one part of the stimulus is indirectly determined by the presence of another part.

In addition to cueing the nature of the consonant, the formant transitions provide information concerning the vowel. Indeed, with a short vowel between certain consonants —as in "did"—there may be no steady portion of the sound-wave with which the vowel might be identified (Lindblom, 1963). Furthermore the precise nature of the sound at the beginning of a syllable will be affected by the consonant which occurs at the end of the syllable (Treon, 1970). We have, then, a situation where a particular part of the sound-wave provides information relevant to the identification of a number of phonemes, and where a particular phoneme is identified from a number of acoustic cues, some imaginary, and some conditional on the presence of other cues which occur at different times in the speech stream.

To all this complication should be added the fact that, since people differ in the sizes of their vocal tracts, there are systematic differences in the sounds that represent the same vowel. Indeed it has been shown that the same sound will be heard as two different words—say, "bit" or "bet"—depending on the nature of the (synthetic) introductory sentence (Ladefoged & Broadbent, 1957).

2. Word Recognition

Having illustrated the lack of direct correspondence between sound and phoneme, we can indicate similar discrepancies between stimulus and response in word recognition. The standard technique for investigating word recognition is to expose the word briefly at low levels of illumination or to play the word in a background of white noise, and to require the subject to respond by saying the word that he believes he has seen or heard. The results referred to below have, in general, been shown for both vision and audition and are not modality-specific. It has been known for some time that the ease of recognition of a word is a function of the frequency of occurrence of the word in the language. Common words such as "table" and "car" are more easily recognized than less common words such as "cable" and "tar". It is now fairly clear that this is because we require less evidence from the senses in order to respond with a high-frequency word, not because we are more efficient in extracting evidence about high-fre-

quency words (Broadbent, 1967; Morton, 1969). Thus we cannot account for the difference between "table" and "cable" by suggesting that the letter "t" is more easily recognized than the letter "c". In addition to responding with the word more readily, it is also fairly clear that we actually perceive high-frequency words more easily, rather than simply guessing them correctly. So we can conclude that the reported sensation is not entirely determined by the clarity of the signal but is influenced by our experience of the linguistic unit that the signal represents. It is not our experience of the signal itself, since visual experience of verbal units affects recognition during subsequent auditory presentation. Neither is it our use of the response which is crucial, since experience in seeing and saying "phrase" does not affect our subsequent visual recognition of "frays" (see Morton, 1964a, for references).

Positive evidence that the unit of processing has an internal (i.e., linguistic–cognitive) rather than an external (stimulus–response) definition is provided by two experiments. The first involved the recall of adverbs derived from adjectives—e.g. "quickly". It was found that the recall of such words was related to the frequency of occurrence of the adjectives from which the words were derived and not to the frequency of the words themselves (Rosenberg, Coyle & Porter, 1966). The second study (Marshall, 1967) involved the recognition of words with two distinct meanings, e.g. "bark" (of the tree or of the dog). The word occurs approximately an equal number of times in the two senses. If recognition is a function of the amount of experience with the linguistic units rather than of the word per se, then the threshold should be that appropriate for a word of half the frequency of occurrence of the stimulus unit. This was found to be the case—a good example of the influence of the inherent (i.e., non-stimulus) properties of the signal.

Recognition can also be facilitated by the use of context, it being well established that words are more easily and more rapidly recognized in sentences than in isolation. The reason for this does not seem to be just that we can guess the identity of a word more easily when we are given a context, but rather that the information from the context can supplement that from the stimulus, the two being treated as equivalent (Morton, 1969). As speakers, we are often sensitive to the requirements of our listeners in this respect. In the extreme we tend to speak slowly and clearly to children and to adults who do not know our language fluently. In this way we give them enough information in the sound-wave to identify each separate word. When talking to fluent listeners, however, we speed up and do not articulate the words as carefully, knowing that the listener can use the context to supplement the signal. In this way we can maximize our rate of talking. More particularly, it will often be the case that a word which is heard clearly in a context will become quite unintelligible when removed electronically and played in isolation.

Lieberman (1963) has shown that the extent to which we articulate a word in fluent speech is a function of its predictability. He asked people to read out sentences, extracted certain words from the recordings, and then played these isolated words to a group of listeners. The clarity of the words was indicated by the number of listeners who identified them correctly. Thus the word "nine" taken from a recording of the sentence "A stitch in time saves nine" was recognized only 50% of the time, while the same word from the sentence "The number that you will hear is nine" was recognized

90% of the time. As "nine" was redundant in the context of the former sentence it was not spoken as clearly. Similar effects occurred for words in the middle of sentences. In the sentence "Neither a borrower nor a lender be", "borrower" and "lender" were both poorly articulated. When the words were isolated from the sentence, "borrower" could be identified only 45% and "lender" 10% of the time; but when the same words were taken from sentences with the frame "The —s were all imprisoned", the figures were 80% and 40% respectively. This experiment makes the further point that we do not just rely on the preceding context to help with the identification of a single word but use the following context as well, if necessary postponing final identification of some words until the rest of the sentence provides information that resolves any ambiguity. Speech recognition is not a simple word-by-word process.

In one sense, then, our recognition of language is in part illusory. When we listen to fluent speech, or read, we are not simply taking in information from the stimulus. Rather, we bring to bear the whole of our knowledge of language in interpreting the sounds or the marks on a page. The consequence of such interaction is that we may make errors in recognition. Thus, when subjects were asked to read aloud they made errors which indicated that, between seeing and saying, the material had undergone a certain amount of syntactic and semantic analysis. What was spoken was a result of a re-synthesis of the material. The correct categories appeared, but the detail was in error—such as "Saturday" for "Sunday", "morning" for "evening", and "the men took off their overcoats" for "the man took off his overcoat"—in which an error becomes extended, unnoticed, over the range of a whole clause (Morton, 1964b).

It is being suggested, then, that when we are presented with language materials we rapidly extract the meaning and structure and ignore the precise form of the stimulus. Such a process can be demonstrated dramatically by presenting people with sentences which deviate from the subject's own grammar. Labov, Cohen, Robins & Lewis (1968) have pointed to several examples of this from their work with New York adolescent negroes. A sentence such as "Nobody ever knew" would be repeated back as "Nobody never knew", since in Non-standard Negro English (NNE) there is a rule requiring "negative concord". Errors of this kind are made in spite of the fact that the youth wins money if he gets the sentence right. Labov *et al.* (1968) also quote the following exchange:

Interviewer: I asked Boot whether he knew how to play basketball.
Subject: I asked Boot did he know how to play basketball.
Interviewer: I said, "I asked Boot *whether* he knew how to play basketball", but you said "I asked Boot *did* he know". Try again. I asked Boot whether he knew how to play basketball.
Subject: I asked Boot ... *whether* ... did he know how to play basketball.

In NNE "whether" does not function to indicate a reported question and would be without a place in the youth's speech. So it is not noted in the sentence he hears, in which the reality is the meaning not the sound. Errors of this kind occur when the sentence to be repeated is well within the memory span of the subject, and with younger children are indicative of the stage of development of their grammar.

3. The Structure of Linguistic Knowledge

We have seen a number of illustrations of the way in which our knowledge interacts with linguistic stimuli that we are recognizing, but have not yet considered the form of this knowledge, especially in relation to the structure of language. The importance of the work of Chomsky (1957) was that he revived the notion that a grammar should describe not just the sentences that had occurred but all sentences that could occur in a language. This made it in principle possible to test the validity of the grammar as a psychological mechanism. His grammar consists of rules whereby sentences may be generated in much the same way that the axioms of geometry generate theorems. The rules are made up of two parts, termed phrase-structure rules and transformational rules (see Lyons, 1968, 1970a, 1970b, for more complete summaries and some alternatives).

The phrase-structure rules account for certain kinds of knowledge about sentences. Take the sentence:

i. The brush broke the window.

We know that the two occurrences of "the" are identical (unlike the two occurrences of "will" in "Will you will your money to me?"), that "the brush" and "the window" go together in a way that "brush broke" does not, that "brush" serves as a noun (whereas in "Did Mary brush with the law?" it serves as a verb), and that "the brush" is the subject of the sentence and "broke the window" is the predicate. Such knowledge can be characterized by representing the sentence as a tree structure:
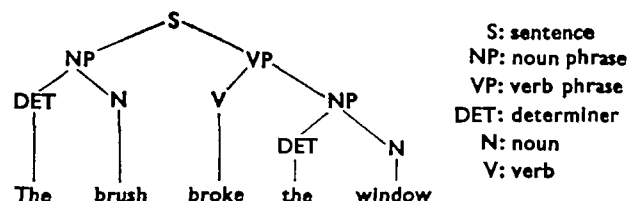
The phrase-structure rules operate to produce successive levels of the tree, with the same rules governing the expansion of NP, for example, wherever it occurs. Chomsky (1957) argued that phrase-structure rules are inadequate by themselves to represent the facts of relations between sentences. He proposed the use of transformational rules. With such rules we can express the relations between (i) and (ii)–(iv):

ii. The window was broken by the brush. (passive)
iii. Did the brush break the window? (question)
iv. Wasn't the window broken by the brush? (passive, negative question)

These and similar sentences would all be produced from the same underlying form or deep structure as (i), since they all express aspects of the same idea. They would differ in the number and type of transformational rules that had been applied.

In some cases, knowledge of the structure of a sentence is almost essential if we are to make a word unambiguous. All the words in (v) are ambiguous except "the":

v. Watch the light one train.

Since only certain combinations of noun, verb, adjective and adverb occur in English sentences, the meaning is easily obtained, but it is likely that knowledge of sentence structure plays some role.

It is clear that syntactic information increases the coherence of a string of items, even if the items themselves are nonsense.

197

Jabberwocky such as (vi) can not only be parsed with ease but is more easily remembered than the same nonsense syllables without their endings (Epstein, 1962):

vi. The yigs wum vumly rixing hum in jegest miv.

It was believed at one time that such differences as the number of transformations would be an index of the difficulty of processing sentences. In addition it was thought possible that when we are asked to remember a sentence we factorize it into its underlying form—the deep structure—plus some tags indicating the grammatical transformations undergone by the sentence. Work by Mehler (1963) indicated that errors in recall of the underlying form of the sentence were independent of errors in the transformations applied, and Savin & Perchonock (1965) produced data which could be interpreted as showing that each additional transformation involved an extra memory load. However, there were other results which demonstrated that sentence processing was not always a simple function of linguistic complexity (see Fodor & Garrett, 1966). This could be because the linguistic theory was incorrect, because its application to behaviour was inappropriate, or because the particular task encouraged different strategies in the subjects.

The last-named possibility is supported by more recent work which has shown that phrase-structure complexity, as well as transformational complexity, exerts some influence on recall (Wright, 1969) and that subjects often remember words and phrases from sentences, rather than always deriving the underlying form and remembering the transformational tags (Wright, 1968).

Some evidence on the sequence of processing during sentence recognition has been obtained using pairs of sentences such as (vii) and (viii):

vii. The reporters (who were watching George) drove to the car.
viii. (To catch the reporters) George drove to the car.

These sentences have clause structures indicated by the brackets. Tape recordings were played to subjects with a sentence in one ear and a click in the other ear. The click was timed to occur in the middle of the word "George". The subjects had to write down the sentences and indicate where they thought the click had occurred. There was a reliable tendency for the subjects to mark the click as having occurred at the clause boundary, that is before "George" in (viii) and after "George" in (vii). This would indicate that the clause forms a natural unit of processing (Fodor & Bever, 1965). Further experiments have indicated that it is the deep, logical or meaning structure of the sentence that is crucial rather than just the phrase structure (Bever, 1971).

In addition to showing the relations between sentences, transformational rules can also give an account of the different underlying structures of sentences with similar surface structures. In (ix) it is Tom who is doing the teasing, whereas in (x) Tom is being teased:

ix. Tom is eager to tease.
x. Tom is easy to tease.
ix(a). Tom is eager that Tom teases someone.
x(a). It is easy that someone teases Tom.

Sentence (x) can thus be regarded as a passive sentence derived from a form such as (x(a)), in which the role of Tom is apparent

in the structure when contrasted with (ix(a)). When subjects are asked to remember sentences like (ix) and (x), it is found that the noun (i.e., Tom) is a more efficient cue to the recall of sentences such as (ix)—where it is the logical subject—than of (x). On the other hand the adjective is a better cue in (x); that is, "easy", which modifies the rest of sentence (x), is a better cue than "eager", which just modifies the noun in sentence (ix) (Blumenthal & Boakes, 1967). Another demonstration of the influence of the logical structure of a sentence involved pairs such as (xi) and (xii):

xi. The child was warmed by the stove.
xii. The child was warm by the stove.

In (xi) "the stove" is the logical subject of the passive sentence. In (xii) "the stove" is part of an adverbial phrase of location. As before, when the word is the logical subject, it is a better cue for the recall of the sentence (Blumenthal, 1967). In this situation then, it is clear that we must be forming a representation at least at the level of deep structure, since there can be no such notion as logical subject before such a level is reached.

These examples illustrate the point that in order to understand a sentence it is necessary to note the relations between words and not just their meanings. Sometimes, however, the relations between the words that carry the informational content are implied by their meanings in any possible sentence. For instance, if "engine", "repaired" and "mechanic" are the only content words in a sentence, their meanings determine the relations between them (since mechanics are not repaired by engines), whatever other words are introduced to complete an active or passive sentence, of whatever tense, with definite or indefinite articles. The meanings of "boy", "girl" and "hit", on the other hand, do not imply a unique set of relations —since either "boy" or "girl" may be the aggressor. This informational analysis finds some support in an experiment of Slobin (1966), who reported that children appear to have special difficulty with reversible passives, such as "The boy was hit by the girl", in deciding whether a particular active or passive statement is true or false of a given picture. There is some evidence that if subjects are not expecting their memory to be tested then even the deep grammatical relations in a sentence are forgotten. Such evidence was provided by Johnson-Laird & Stevenson (1970), who asked subjects to listen to a six-sentence passage of which (xiii(a)) was the fourth sentence. Then they were asked to select the sentence in the passage from a list of eight, including (xiii(a–d)) and four others with "John" and "the duchess" exchanged:

xiii(a). John liked the painting and bought it from the duchess.
xiii(b). John liked the painting and the duchess sold it to him.
xiii(c). The painting pleased John and he bought it from the duchess.
xiii(d). The painting pleased John and the duchess sold it to him.

The point about these sentences is that their deep structure cannot be reconstructed from the meaning; and, of those subjects unaware that their memory would be tested, ten out of twelve were incorrect in their choice compared with five out of twelve subjects who knew they would be tested but did not

know which sentence in the passage was the critical one. Similar results have been found by other workers (see Johnson-Laird, 1970).

The studies described in this section show that the demands of particular tasks modify the influences of linguistic structures on performance. They imply that an account of sentence meaning is likely to have to recognize the significance of a

conglomeration of ad-hoc devices, in addition to the kinds of relations that have been outlined by linguists; and it remains possible that Chomsky's notion of deep structure may have to be replaced by something much closer to the meaning of a sentence. The ground is being cleared, but, as was implied at the beginning of this paper, the path from sound to meaning remains a very indirect one.

REFERENCES

Bever, T. G. (1971) In: Morton, J., ed. *Biological and social factors in psycholinguistics*, p. 158. Logos Press, London

Blumenthal, A. L. (1967) *J. verb. Learn. verb. Behav.* 6, 203

Blumenthal, A. L. & Boakes, R. (1967) *J. verb. Learn. verb. Behav.* 6, 674

Broadbent, D. E. (1967) *Psychol. Rev.* 74, 1

Chomsky, N. (1957) *Syntactic structures.* Mouton, 's-Gravenhage

Epstein, W. (1962) *Am. J. Psychol.* 75, 121

Fodor, J. A. & Bever, T. G. (1965) *J. verb. Learn. verb. Behav.* 4, 414

Fodor, J. [A.] & Garrett, M. (1966) In: Lyons, J. & Wales, R. J., ed. *Psycholinguistic papers*, p. 135. (Proceedings of the 1966 Edinburgh conference.) University Press, Edinburgh

Johnson-Laird, P. N. (1970) In: Lyons, J., ed. *New horizons in linguistics*, p. 261. Penguin Books, Harmondsworth

Johnson-Laird, P. N. & Stevenson, R. (1970) *Nature, Lond.* 227, 412

Labov, W., Cohen, P., Robins, C. & Lewis, J. (1968) *A study of the non-standard English of Negro and Puerto Rican speakers in New York City.* (Cooperative Research Project no. 3288.) Vol. I: *Phonological and grammatical analysis.* Office of Education, U.S. Department of Health, Education and Welfare. Government Printing Office, Washington, D.C.

Ladefoged, P. & Broadbent, D. E. (1957) *J. acoust. Soc. Am.* 29, 98

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) *Psychol. Rev.* 74, 431

Lieberman, P. (1963) *Lang. Speech,* 6, 172

Lindblom, B. (1963) *J. acoust. Soc. Am.* 35, 1773

Lyons, J. (1968) *Introduction to theoretical linguistics.* University Press, Cambridge

Lyons, J. (1970a) *Chomsky.* Fontana, London

Lyons, J., ed. (1970b) *New horizons in linguistics.* Penguin Books, Harmondsworth

Marshall, J. C. (1967) *Studies of word recognition and grammatical organisation: some psychological concomitants of linguistic structure.* (Thesis for Ph.D. degree) University of Reading

Mehler, J. (1963) *J. verb. Learn. verb. Behav.* 2, 346

Morton, J. (1964a) *Int. Audiol.* 3, 216; reprinted in Oldfield, R. C. & Marshall, J. C., ed. (1968) *Language: selected readings*, p. 147. Penguin Books, Harmondsworth

Morton, J. (1964b) *Lang. Speech,* 7, 40

Morton, J. (1969) *Psychol. Rev.* 76, 165

Rosenberg, S., Coyle, P. J. & Porter, W. L. (1966) *J. verb. Learn. verb. Behav.* 5, 75

Savin, H. B. & Perchonock, E. (1965) *J. verb. Learn. verb. Behav.* 4, 348

Skinner, B. F. (1957) *Verbal behavior*, p. 3. Appleton-Century-Crofts, New York

Slobin, D. I. (1966) *J. verb. Learn. verb. Behav.* 5, 219

Treon, M. A. (1970) *Lang. Speech,* 13, 54

Wright, P. (1968) *Q. Jl exp. Psychol.* 20, 265

Wright, P. (1969) *Br. J. Psychol.* 60, 63

.

199