4 The Sapir–Whorf hypothesis

The subject of this chapter is not so much a geographically or chronologically distinguishable school of thinkers, as an idea which has held a perennial fascination for linguists of diverse schools, and indeed for very many people who have never been students of language in any formal sense. This idea – that a man's language moulds his perception of reality, or that the world a man inhabits is a linguistic construct – although in one form or another a very old one, has become associated with the names of the Americans Edward Sapir (1884–1939) and Benjamin Lee Whorf (1897–1941), and more particularly with the latter.

The work of these writers might well have been treated in the last chapter, since it fell squarely within the tradition initiated by Boas. I have chosen to discuss Sapir and Whorf in a separate chapter, because the aspect of their work which we shall examine represents a rather special development within the Descriptivist school, and one which conflicted fairly sharply with the thought of many other members of that school. Sapir and Whorf fully shared the relativism of Boas and his Descriptivist successors, with its emphasis on the alienness of exotic languages, while never being influenced by the behaviourism (in either 'good' or 'bad' senses) of Bloomfield. (Behaviourism was an element which Bloomfield imported into the Descriptivist tradition rather than finding it already there – Boas, and indeed Bloomfield himself in his early writing, were happy to discuss meanings and spent little time worrying about the logical status of linguists' data. But Bloomfield succeeded in taking most of his colleagues with him in his conversion to behaviourism, which is why I say that there was a conflict between the ideas summarized as the 'Sapir–Whorf hypothesis' and the ideas of other Descriptivists.)

Sapir studied languages of the Pacific coast of North America,

and began his career in charge of anthropological research at the Canadian National Museum; in 1925 he moved to the University of Chicago, and in 1931 to Yale. Much of his work was quite comparable to that of other Descriptivist linguists, though he differed from the behaviourists in stressing that patterns revealed by linguistic analysis were patterns in speakers' minds (it is significant that the collection of his papers published in 1949 bears the title *Selected Writings in Language, Culture, and Personality*), and Sapir took it for granted that if one wants to know how a language is structured for its speakers it is appropriate to ask them.[1] Sapir's independence of his American colleagues' assumptions is particularly obvious in his notion of linguistic 'drift': behind the more-or-less random fluctuations which make up the detailed history of any language, Sapir thought, there was a long-term tendency for that language to modify itself in some particular direction, as the coming and going of waves on a beach masks a steady long-term tidal movement (Sapir 1921, ch. 7). This idea comes very close to implying that a language has a life of its own in some more than metaphorical sense, and it would clearly have been anathema to a methodological individualist such as Bloomfield.

On the issue with which this chapter is concerned, Sapir was by no means single-minded. The occurrence of his name in the term 'Sapir–Whorf hypothesis' is perhaps due more to the fact that Whorf took his general approach to linguistics from Sapir than to Sapir's being one of the most active proponents of that hypothesis. (The term was introduced by J. B. Carroll (Whorf 1956, p. 27).) In his popular book *Language*, indeed, Sapir suggests that differences between languages are merely differences in modes of expressing a common range of experiences, rather than corresponding to differences in the experiences themselves (Sapir 1921, p. 218). Later, though, Sapir changed his mind. Consider, for example, the following passages:

> Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society. It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that *the 'real*

world' is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached. [1929, p. 209; my italics]

Language . . . not only refers to experience largely acquired without its help but actually *defines experience for us* by reason of its formal completeness and because of our unconscious projection of its implicit expectations into the field of experience . . . . Such categories as number, gender, case, tense, . . . are not so much discovered in experience as imposed upon it because of the tyrannical hold that linguistic form has upon our orientation in the world. [1931; my italics]

These remarks might be interpreted as mere truisms, but if taken literally they are strong statements. The special contribution of Whorf was, by means of detailed analysis of certain American Indian languages, to make as convincing a case as has ever been made for believing that we must acknowledge the view expressed by Sapir as true in a quite radical, untrivial sense.

Benjamin Lee Whorf, a descendant of seventeenth-century English emigrants to Massachusetts, was in his scholarly work an outstanding example of the brilliant amateur. After taking a degree in chemical engineering he began a successful career as a fire-prevention inspector with an insurance company in Hartford, Connecticut, and despite several offers of academic posts he continued to work for the same company until his death at the age of 44. (Whorf learned lessons from his professional work which encouraged his belief that world-view is moulded by language. In analysing a large number of reports of how fires had started, Whorf tells us (1941a, p. 135), he began by assuming that only physical factors would be relevant but came to realize that language often played an important role: for instance, people behaved cautiously near what they categorized as 'full petrol drums' but carelessly near 'empty petrol drums', although the 'empty' drums contained explosive petrol vapour and were thus even more dangerous than the full ones.) Whorf's linguistic interests were originally rather diverse; when in 1931 Sapir moved to Yale University, only thirty-odd miles from Hartford, Whorf became a regular collaborator of his and began to focus his attention mainly on Hopi, a language of

Arizona. Much of Whorf's writing discusses the special, very un-European world-view which he believed to be implied by various features of Hopi grammar.

Whorf makes the point that only certain grammatical categories in any language are marked overtly, as, for example, the distinction between present and past tense is marked in every finite verb which occurs in English. There exist also numerous 'covert' categories, or 'cryptotypes' as Whorf sometimes calls them. For instance, in English the names of countries and towns form a 'cryptotype' because, although they outwardly resemble other nouns, they cannot be reduced to pronouns after the prepositions *in, at, to, from* (Whorf 1945, p. 92). Thus one can say *I live in it* when 'it' refers back to a phrase such as *that house* or *the basement*, but not when it refers to *Kendal* or *Bulgaria* – even though *I live in Kendal, I live in Bulgaria* are perfectly correct. Whorf felt that such covert categories were more telling than the overt categories of a language in establishing the world-view of its speakers, on the ground that the use of overt markers may be merely learned by rote but 'cryptotypes' can be manipulated consistently only if the categorization which they imply is real for the speaker. (If all country-names and town-names ended in some special suffix, say *-ia*, then an Englishman could simply remember 'nouns ending in *-ia* may not pronominalize after a preposition', but since they in fact have no special form we must think of them as a semantic class.) In Hopi rain-prayers, it seems that clouds are spoken of as if they were alive. Whorf points out that from this alone one cannot know whether the usage 'is some metaphor or special religious or ceremonial figure of speech', or whether the Hopi actually believe that clouds are living beings. However, the distinction between animate and inanimate exists as a covert category in Hopi. Any noun used to refer to a living being is pluralized in a special way (even when the noun is not basically animate, so that, for example, the Rolling Stones in Hopi would take the animate plural of 'stone'); and the word for 'cloud' is invariably pluralized in the animate way, which demonstrates that the Hopi do indeed believe clouds to be alive (Whorf 1956, p. 79).[2]

Although this neatly illustrates Whorf's point about the importance of covert categories, it is not a particularly good example of the differences Whorf claims to exist between Hopi

and European world-views: in this case the *categories* animate/inanimate are perfectly normal for a European, and the only question concerns the status of clouds with respect to these categories. (We shall consider a better example of Whorf's thesis about linguistic diversity shortly.) Nevertheless, even here it is possible to take a sceptical stance. Thus, suppose that we encounter another tribe in which sex is a 'covert category', so that, say, all nouns referring to females evoke special suffixes in words modifying them; and suppose further that many words for inanimate objects, such as 'stone', 'water', 'moon', belong to the female cryptotype, while others, such as 'iron', 'fire', 'sun', behave like the words for males. Clearly Whorf would have to conclude that this tribe holds some sort of animistic view of Nature, according to which everything that exists is alive and has a sex. But there is such a tribe: they live just across the Channel from Dover, and if there is one thing the French are not it is surely animists. Whorf did not in fact apply his notions to differences between the familiar European languages; he felt that these all presupposed the same world-view because of the long period in which Europe had shared a common culture, and he referred to them collectively as 'Standard Average European'. It is perhaps appropriate to be cautious, at least, in accepting a theory which says that certain communities see the world in ways startlingly different from ours, but which is illustrated almost wholly by reference to primitive tribes about whose beliefs we have little independent evidence. The non-European language with which the present writer is best acquainted is Chinese; although traditional Chinese ideas about the world differ greatly from European ideas, the two intellectual systems do not seem to possess quite the same quality of mutual incommensurability that Whorf alleges to occur with Hopi *vis-à-vis* 'Standard Average European'. One cannot help wondering whether this may be because Chinese civilization, although, like that of the Hopi, quite independent of Europe, has been articulate enough to refute the flights of fancy in which a Whorf might be inclined to indulge on the basis of formal characteristics of Chinese grammar.

In fact, the various contrasts in world-view for which Whorf argues differ greatly in the extent to which they are surprising or controversial. Boas had already made the point that, for instance, where English has the one word *snow* Eskimo has

separate basic roots for snow falling, snow on the ground, drifting snow, and so forth; at this relatively concrete level disparities between the conceptual schemes of different languages are fairly familiar, and there is no doubt that they influence perception – it can be shown that people's perceptions of their surroundings are modified by the conceptual categories their language happens to provide (Lenneberg and Roberts 1956, p. 31; cf. Herman *et al.* 1957; Hanson 1958). Whorf discusses cases of this kind, but they are not what he is primarily interested in. 'What surprises most', Whorf rightly says, 'is to find that various grand generalizations of the Western world, such as time, velocity and matter, are not essential to the construction of a consistent picture of the universe' (1940, p. 216). Hopi, in particular, 'may be called a timeless language': the language does not recognize time as a linear dimension which can be measured and divided into units like spatial dimensions, so that for instance Hopi never borrows spatial terms to refer to temporal phenomena in the way so common in European languages (*before the door ~ before noon, between London and Brighton ~ between 9 and 10 a.m., in the box ~ in the morning*), nor does Hopi permit phrases such as *five days* since daytime is not a thing like an apple of which one can have one or several. Furthermore, Hopi verbs do not have tenses comparable to those of European languages. And since there is no concept of time, there can be no concept of speed, which is the ratio of distance to time: Hopi has no word for 'fast', and their nearest equivalent for 'He runs fast' would translate more literally as something like *He very runs*. If the Hopi rather than Europeans had developed sophisticated scientific theories, Whorf suggests, modern physics would be very different from what it is, though it might be equally self-consistent and satisfactory.

One objection to this interpretation of Hopi thought, voiced, for example, by Max Black (1959), is that Whorf's claim is untestable and therefore vacuous. It might be that the Hopi have much the same concept of time as we have, but simply use somewhat exotic turns of phrase in talking about matters of time – *He very runs* is just their way of saying 'He runs fast', and they mean by their sentence just what we mean by ours. After all, an Englishman calls a huntsman's coat 'pink' but that does not imply that he sees it as other than red. Whorf admits that

'the Hopi language is capable of accounting for and describing correctly, in a pragmatic or operational sense, all observable phenomena of the universe' (1956, p. 58); could any evidence then force us to conclude that the difference between Hopi and English ways of talking about time is more than a difference in formal modes of expressing a common range of ideas? (Philosophers will recognize that Willard Quine in *Word and Object* (1960) and subsequent writings has argued, essentially, for a negative answer to questions of this kind.)

There may be a reply to this objection (as it relates to Whorf's claims – not to the more general point made by Quine). In the first place, it may be that there are indeed observable aspects of Hopi behaviour which correlate with their 'timeless' outlook on life; cf. Whorf (1941a, pp. 148, 153). I have read that Indians on reservations in the US Southwest (unfortunately I do not remember whether this was said specifically of Hopi) have difficulty in holding jobs in the white man's economy because they cannot get into the habit of catching commuter buses and generally keeping to a timetable, and this might be evidence in favour of Whorf's ideas. True, the sceptic could point out that some individual Englishmen have similar problems, and we do not usually ascribe this to causes so lofty as a non-standard philosophy of time. But if the sceptic argues that the reason for American Indians missing buses is mere idleness or fecklessness rather than a special view of time, it would perhaps seem an awkward coincidence that the communities in which this idleness is unusually widespread are also communities which speak languages that treat time in an odd way.

Furthermore, even if independent evidence did little to corroborate Whorf's claims, I am not sure that Black's objection need be fatal. It is perhaps wrong to suppose, because of the word 'hypothesis' in the standard name for Whorf's idea, that it is to be interpreted as a scientific theory which makes testable predictions about observable data. It might be more appropriate to interpret Whorf as giving a philosophical account of alternative conceptual frameworks, which could not be confirmed or refuted by facts observed from within any one of those frameworks. (To give a parallel: we can contrast the mediaeval system of arguments from authority with the modern 'scientific method' of proposing and testing falsifiable

hypotheses, but we cannot usefully cite evidence showing that the latter method of discovering truth is superior to the former, since the question whether it is appropriate to cite evidence for one's beliefs is exactly what is at stake.) Ludwig Wittgenstein in his later writings argued a view very similar to Whorf's (though without Whorf's knowledge of exotic languages) about the interdependence of world-view and language, and Wittgenstein was quite clear that he could only ask his readers to 'see' that his account was correct, he could not demonstrate it to be so; ironically, while Black attacks Whorf's 'hypothesis' for untestability he strongly supports Wittgenstein's frankly untestable philosophy.

Where Whorf (although not Wittgenstein) does fall down badly is in his apparent inability to allow for the radical changes of world-view which occur within a given linguistic community. Since we are discussing time and space, the obvious example to quote here is Albert Einstein. Einstein's new account of the 'grand generalizations' of physics seems fully as alien, from the standpoint of received views, as the Hopi approach; yet Einstein spoke a Standard Average European language. The history of science over several centuries has been a history of repeated radical changes of world-view, almost all of which occurred within the Standard Average European linguistic framework. Whorf (1941a, p. 153) supposes that Newton's physics was given to him ready-made by his language, but this idea that Newtonian physics is just common sense rendered explicit is an illusion deriving from the long period in which Newtonian physics has been accepted as true. As Black points out (1959, p. 254), Descartes, although also a 'Standard Average European' speaker, had worked out a very different structure of spatial concepts from those later evolved by Newton, and Newton's account was preferred not because it conformed better to men's commonsense ideas but because it turned out to be closer to the truth. Rather than saying that if the Hopi had developed physics then physics would look very different, it might be more appropriate to say that if the Hopi had developed physics then the Hopi world-view would have changed (and, by the same token, presumably the Hopi language is unsuitable for discussing bus timetables because the Hopi have not had much to do with buses, rather than *vice versa*).

Of course it is true that each of us accepts many inherited presuppositions, and such presuppositions may well be reflected in our language; but none of our inherited prejudices are sacrosanct, and human thought consists of a constant process of individuals questioning received presuppositions and replacing them by new and better ideas, which then become later generations' 'commonsense' until another individual has a still better idea. As the German philosopher J. G. Hamann wrote in 1760, 'a mind which thinks at its own expense will always interfere with language' (quoted by Cohen 1962, p. 10). Certainly the language of a community and the thought of individual members of the community each influence the other, but what ultimately counts is the individual's influence on the language; the influence of language on individual is a purely negative matter of the individual's failure to examine critically all the ideas of various earlier individuals. Sapir and Whorf write as if language exerted a positive influence, and one with far more power than the reverse influence: Sapir writes of individuals being 'at the mercy of' their language, which exerts a 'tyrannical hold' over their mind (cf. pages 82–3 above), Whorf writes of speakers being parties to an 'absolutely obligatory' agreement to conceptualize the world in a certain way (1940, pp. 213–14). It seems to me that this 'tyranny' is of the same order as the 'tyranny' to which my body is subjected early on a Monday morning by my bed. Sapir's and Whorf's account of the situation may not be too inaccurate in practice, but that is only because many people are mentally very lazy. To quote Imre Lakatos (1976, p. 93 n.): 'Science teaches us not to respect any given conceptual–linguistic framework lest it should turn into a conceptual prison – language analysts have a vested interest in at least slowing down this process [of conceptual change].'

There is a further problem that arises when Whorf's hypothesis is taken in its most radical interpretation: it may actually be self-contradictory. The most fundamental component of any semantic structure, deeper even than the 'grand generalizations' of physics, is its logical apparatus – in English, the use of words such as *not, if, all* and so forth. One might take the Whorf hypothesis to mean that even logic is relative to language, so that, say, if Aristotle had been a Hopi then modern logic as well as modern physics would have developed quite differently. There are hints in Whorf's writings (e.g. 1941b,

p. 241) that he meant to go as far as this, and other linguists have put the point plainly: see for example Sommerfelt (1938, p. 9), Benveniste (1958), Hjelmslev (1963, p. 121). If these writers mean only that formal characteristics of language have influenced the explicit systems of logic which philosophers have devised in their fallible attempts to *describe* publicly the patterns of our (largely unconscious) thought-processes, they are no doubt right. If, though, they mean that those thought-processes in their logical aspects are themselves a function of our language, then their notion must be rejected on *a priori* grounds.

To see this, let us turn to a predecessor of Sapir and Whorf who argued for the view that I regard as untenable perhaps more fully than anyone else has done, namely the French anthropologist Lucien Lévy-Bruhl (1857–1939). Lévy-Bruhl's view of the relation between language and thought was in general similar to Whorf's (Lévy-Bruhl 1910, ch. 4), except that he did not share Whorf's relativism: rather than thinking of Standard Average European as one among a diverse range of alternative conceptual frameworks, Lévy-Bruhl believed that the thought-patterns of all primitive peoples were similar as contrasted with the thought-patterns of civilized men. Lévy-Bruhl did not suggest that the distinction between savages and civilized men was a sharp one, but for him different men's minds occupied different points on a single scale. The most important aspect of the distinction in mental type was a matter of logic: according to Lévy-Bruhl, the primitive mind does not acknowledge the law of non-contradiction.[3] That is, whereas a civilized man regards any statement of the form 'P and not P' as self-evidently false, a primitive man will regard many such statements as true and will see no difficulty therein. (It is true that all of us make statements such as 'I want to go and I don't want to go', but these are intended to be understood in ways which make them non-contradictory, e.g. as 'There are reasons why I should like to go and other reasons why I should not'. We succeed in interpreting such statements correctly just because we do recognize the law of non-contradiction and therefore know that they cannot mean what they appear to mean. Lévy-Bruhl argues that primitive men, on the other hand, believe contradictions in which each half is understood literally and unequivocally.)

To quote one of many pieces of evidence which Lévy-Bruhl

cites in favour of this claim, the Bororó of northern Brazil are said by Karl von den Steinen to believe that they are red parakeets (although, one must presumably add to get the 'not P' side of the contradiction, they can obviously see that they are not red parakeets):

> This does not merely mean that after their death they become parakeets, nor that parakeets are metamorphosed Bororó, and must be treated as such. It is something quite different. 'The Bororó', says von den Steinen, who was reluctant to believe it, but finally had to give in to their explicit affirmations, 'give one rigidly to understand that they *are at the present time* parakeets . . . .' [Lévy-Bruhl 1910, p. 77, quoting – somewhat loosely, it should be said – von den Steinen 1894, p. 352]

I find Lévy-Bruhl's explanation for findings such as von den Steinen's a quite unsatisfactory one, for one thing because it could so easily be turned round against the 'civilized' mentality. One can well imagine a Bororó who had visited Europe announcing to a meeting of the Bororó Anthropological Association, with a superior smile, that the wise men of that region claim with every appearance of sincerity that coal and diamonds are the same substance: 'this does not merely mean that they have a method of making diamonds out of coal, or anything of that sort; the white men give one rigidly to understand that a lump of coal consists of the same stuff as a diamond *at the present time*, but I found their attempts to explain the nature of this identity quite impossible to follow (sniggers in the auditorium); clearly the whites do not recognize the law of non-contradiction.' What distinguishes the Bororó, as described by von den Steinen, from the European is surely not a matter of logic but of beliefs about fairly abstract matters of fact: each community holds certain sophisticated theories which are only very indirectly connected with observable reality, and these theories cannot be merely *translated* but must be *taught* at length to members of the other community, just as they must be taught to young members of the community which has evolved them. We have no more right to call the Bororó mentality 'pre-logical' because of their theory about parakeets than they have to call us pre-logical because of Western chemistry or the doctrine of the Trinity. (It *may* be that our doctrines are truer than theirs, but a false belief is not necessarily an illogical belief.) Granted that no man is what Bertrand Russell calls a 'logical saint' – none of us works out all the innumerable

implications of his beliefs and weeds out all the sources of contradiction which his beliefs contain – still Lévy-Bruhl gives us no reason to think that savages are greater logical sinners than we.

It is difficult to know exactly what Lévy-Bruhl means by calling savages pre-logical – he hedges his bets to some extent, and in later writings he abandons the notion of a 'pre-logical mentality' entirely in response to criticisms similar to mine. But suppose I am right in interpreting him as saying that savages will typically believe certain statements which translate into English as 'P and not P' (for some statement P). I have shown that the kind of evidence Lévy-Bruhl gives does not *require* us to accept this: let me show that no conceivable evidence could even *allow* us to accept it.

Let us say that the savage shows signs of assenting to a sentence which in his own language runs *P ka bu P*, and an anthropologist claims that *P* translates into some simple English statement, that *ka* means 'and', and that *bu* means 'not'. How does the anthropologist know how to do the translation? For some words the translation can be worked out by observation of the external world: if the savage points to a parakeet and says *Arara!* it is likely (though far from certain) that *arara* means 'parakeet'. In the case of more abstract words, the evidence of observation is less helpful: if the savage uses the word *vekti* to explain why he hands over some of his goods to a man who comes to the door, we may at first suppose that *vekti* means 'tax', but when that assumption forces us to translate a subsequently heard remark as 'Nobody is required to pay tax' or as 'It gives one a good feeling to pay tax' we are likely to change our mind and translate *vekti* as 'charity'. In other words, an important part of what makes a system of translation 'correct' is that it translates sentences which speakers of the source-language regard as true into truths of the target-language, that it translates falsehoods into falsehoods, nonsense into nonsense, tautologies into tautologies, and so on. We cannot expect perfect matching: it might be that the French hold the sentence *La Concorde, c'est l'avion de l'avenir* to state a truth while the English regard 'Concorde is the aeroplane of the future' as false, and this is clearly not enough to show that the two sentences mean different things. But, if the *majority* of translations which are generated by the system we learn at

school for turning French into English turned out to have truth-values which contrasted with those of the original sentences, then we would have to conclude that the traditional system of translation embodied a serious misunderstanding of the French language. (Clearly this is not in fact so: cases of the 'Concorde' type are a small, though interesting, minority among all accepted translations between the two languages.)

Now logical words like 'not' and 'and' are words whose meanings are ascertained exclusively from 'internal' evidence of this sort, rather than by observation of the external world (one can show someone a parakeet, but one can hardly show them 'and'). Furthermore, unlike the case of 'charity', for 'not' and 'and' the relevant internal evidence is very simple and straightforward. To say that a word means 'not' is to say that the word changes a truth into a falsehood and *vice versa*; to say that a word means 'and' is to say that the complex sentence formed by inserting it between two simpler sentences is true if and only if both the simple sentences are true. (Cases where 'and' links elements which are not sentences, *e.g. John and Mary* . . . , may be ignored here.) From this it follows that to say that some sentence means 'P and not P' is to say that the sentence as a whole cannot be true, irrespective of the meaning of P (since if 'P' is true 'not-P' must be false, and therefore the sentence as a whole must be false; and if 'P' is *not* true then the sentence as a whole is not true).

In other words, evidence that the savage believes *P ka bu P* to be true *is itself the best possible evidence* that that sentence does *not* mean 'P and not P'.

It is senseless to claim that a savage (or anyone else) believes an explicit contradiction, because to believe any proposition entails understanding it, and to understand a contradiction *is* to recognize that it is necessarily false. Probably all of us hold beliefs which lead to contradictions at the end of chains of inference which we have not worked out, but that is a different matter. It is conceivable that one might encounter a community speaking a language which was wholly untranslatable, in the sense that *no* systematic scheme for translating its sentences into sentences of a European language generated more matches between truths and truths, falsehoods and falsehoods, tautologies and tautologies, etc., than would occur if translations were chosen at random. (It is interesting that no such language

has ever been reported, since there is no obvious reason in logic why such a language might not exist.) What is excluded *a priori* is a language for which there is a 'correct' system of translation, but whose speakers disagree with us not just on specific matters of fact but on basic principles of logic.

I do not suppose that von den Steinen or Lévy-Bruhl erred by mistakenly translating some Bororó word as 'not' or 'and'. Much more probably, they correctly translated a standard Bororó assertion as, say, 'We are red parakeets', and mistakenly supposed that observation of their own bodies would force the Bororó (as it forces us) also to believe 'We are not red parakeets' – but Bororó theories are such that this does not follow, just as our theories are such that the sight of a diamond does not force us to the conclusion 'This is not the same substance as coal'. But whatever Lévy-Bruhl meant by calling the primitive mind 'pre-logical', the general point is made. The deeper and more abstract are the aspects of a 'world-view' which are claimed to be a function of language, the more compelling becomes the argument that alleged differences in world-view result from misinterpretation of language. Concepts of space and time are already distant enough from immediate observation to make Whorf's claim very difficult to substantiate; in the case of logical concepts, the argument for mistranslation is certain to succeed. Whorf is on much firmer ground with cases such as the Eskimo's several words for types of snow, because words of that sort are linked relatively closely to observable reality, and the possibility of mistranslation is accordingly remote. But this means that the Sapir–Whorf hypothesis is most plausible where it is relatively trivial.

I say that the hypothesis is 'trivial' insofar as it refers to differences in the categorization imposed by various languages on concrete, observable phenomena, because examples of such differences are familiar to many people and this aspect of the hypothesis was until recently quite uncontroversial. In the case of colour, for instance, it is well known that various languages cut the visible spectrum up in different ways: thus Welsh subsumes our 'blue' and 'green' under a single word *glas*, while Russian uses separate words, *sinij* and *goluboj*, for our 'light blue' and 'dark blue', respectively. H. A. Gleason's popular elementary textbook of linguistics gives the following diagram as its very first illustration of the differences between linguistic

structures (Gleason 1969, p. 4):

Figure 2

English

purple	blue	green	yellow	orange	red

Shona (a language of Rhodesia)

cips^{w}uka	citema	cicena	cips^{w}uka

Bassa (a language of Liberia)

hui	zĩza

SOURCE: H.A. Gleason, *Introduction to Descriptive Linguistics* (New York: Holt, Rinehart & Winston, 1969).

(Note that the Shona system has three, not four, terms – our orange, red, and purple are all *cips^{w}uka*; note also that *citema* covers black as well as blue and blue-green, and *cicena* covers white as well as yellow and some greens.)

Colour is in fact a particularly favourable arena for the Whorf hypothesis, possibly the most favourable of all. It is an immediate property of observed sense-data: to find out whether *vekti* meant 'tax' or 'charity' we had not only to observe but also to investigate speakers' beliefs about *vekti*, but a red patch is a red patch irrespective of the beliefs of the man who sees it. And, among perceptual variables, that of colour is one in which we are physically capable of making a very large number of distinctions (there are estimated to be at least 7,500,000 discriminable shades of colour), so the question how these are grouped into classes in any given language is very far from trivial. Furthermore, physics provides us with a neutral, objective standard against which to compare the terminologies of different languages; and, most important, the world of colour appears to have no natural boundaries – it seems a featureless steppe on which colonists must draw their frontiers where they will, rather than a continent which Nature has already parcelled up by means of mountain ranges and wide rivers. So, if Whorf's hypothesis applies anywhere, it should certainly apply to colour; and linguists have long taken it for granted that it does.

Against this background, two anthropologists of the

University of California at Berkeley, Brent Berlin and Paul Kay, caused a considerable stir in 1969 by publishing a book, *Basic Color Terms*, which argues, on the basis of copious evidence, against linguistic relativism in the very field where it seemed so secure. Berlin and Kay belong to the new movement in linguistics which holds that human languages are all cut to a common pattern which is determined by psychological structuring innate in our species (we shall return to this notion in later chapters). While they have no quarrel with Whorf's claim that the nature of a language and the world-view of its speakers are intimately connected, they object to the other half of the Whorfian hypothesis, namely that language structures (and their associated world-views) are highly diverse. Berlin and Kay obviously recognize that there do exist differences between the colour terminologies of various languages, but they argue that the differences are relatively superficial matters which mask certain deep underlying principles common to the colour terminologies of all languages.

Berlin and Kay begin by investigating the colour terminologies of twenty languages from widely scattered areas of the world, using native-speakers' judgements of how to label various portions of a large standard colour chart. For each language, they set out to establish a class of most-basic colour words, excluding terms for finer shades (e.g. for English *red* is included, *vermilion* rejected because it is a subdivision of *red*); they use several formal clues to help in this, thus a colour term is probably non-basic if it is morphologically complex (e.g. *yellowish, sky-blue*) or borrowed from another language (e.g. *maroon* from French *marron*) or if it also refers to a thing of the relevant colour (e.g. *silver* or *chocolate* – though they are forced to admit exceptions, such as *orange* in English). After analysing the results of this stage of the research, they supplement their data by using the patterns that emerge from the analysis to interpret published descriptions of the colour terms of a further seventy-eight languages for which they had no access to native informants.

In analysing their results, Berlin and Kay first make the very astute remark that previous writers erred in concentrating on the *boundaries* of the domains of various colour terms, while what matters are the *focal points* or 'best examples'. (Concentration on boundaries was clearly encouraged by

Saussure's 'structuralist' approach to semantics – cf. Figure 1 (page 40) – but is not really an essential principle of the Saussurean and Boasian semantic relativism which Berlin and Kay aim to refute.) To continue with our metaphor of the colour terms of a language as rival colonies dividing up a continent, we should think of the colonies not as territories with formal frontiers, but as city-states whose control over the surrounding land diminishes gradually with distance, so that there are many border areas of doubtful allegiance. One knows how hard it is to decide whether certain shades are 'green' or 'blue' – many people would call the official colour of Cambridge University light green; there is much less disagreement about what shade is the 'greenest green' or 'bluest blue'.

Berlin and Kay then investigate the distribution of 'focal colours' as identified by the informants on their standard colour chart, which is a two-dimensional array of 320 samples at forty steps along the hue dimension and eight steps on the tone dimension. (Hue is the perceptual variable corresponding to wavelength, i.e. position in the rainbow spectrum; tone is the lightness or darkness of a colour, thus all hues vary along the tone dimension from white through pale, mid, and dark shades of that hue to black. The distinction in English between *pink* and *red* is mainly a distinction of tone. Berlin and Kay supplement the main 40 × 8 chart with a series of nine neutral greys of various tones.)

When the focal points for various colour terms in various languages are all plotted on a single copy of the chart, they turn out to cluster in certain quite limited areas, rather than being scattered randomly over the whole chart (as Whorf might predict). For instance, one particular sample in the yellow area was chosen as the focal point of a colour-term for eight languages, and its neighbouring samples also scored well, although, since the twenty languages investigated at first hand had 127 basic colour terms between them (ignoring words for black, white and grey) and the colour chart included 320 samples, an 'average' sample would have scored only 0.4. Berlin and Kay therefore identify eleven smallish areas of the chart as 'universal colours' (the phrase is mine, not theirs), corresponding to the English words *red, pink, orange, yellow, brown, green, blue, purple, black, white, grey*.

Not every language investigated codes each of the eleven

'universal colours'; languages with words for all eleven tend to be languages of technologically advanced civilizations, while primitive tribes have far fewer colour-names.[4] But Berlin and Kay go on to show that there is considerable patterning in which of the 'universal colours' are coded in simple systems. The minimal system ('Stage I'), naturally enough, has just two terms whose focal examples are black and white (although, in such languages, 'black' and 'white' cover dark and light shades, respectively, of every hue). None of the twenty systems which Berlin and Kay themselves studied was as simple as this, but they cite reports by colleagues and others of several such languages, mostly spoken in New Guinea. If a language has three colour names, the focal point of the third will be red. (Shona, as described by Gleason – cf. Figure 2, page 95 – is an example of this 'Stage II', three-term system. From Gleason's description it is impossible to tell whether or not his other language, Bassa, fits Berlin and Kay's analysis, since Gleason does not explain how Bassa treats black and white.) A four-term system will have black, white, red and either green or yellow; and a five-term system will have the first three together with green *and* yellow. Only a system with at least six terms will have blue; a seven-term system will add brown; and finally purple, pink, orange, and grey may occur in any combination in languages which also have all seven of the earlier universal colours. Languages with, say, a four-term system of black, white, red, and blue simply do not exist.

All in all, Berlin and Kay appear to have dealt a severe blow to the notion of linguistic relativism. If even this area of semantics manifests such constancies between widely-separated cultures, are there likely to be many areas in which people's world-views really are free to differ?

However, a closer examination of Berlin and Kay's work reveals a number of problems which, taken together, leave one in some doubt as to what, if anything, they have demonstrated.

In the first place, Berlin and Kay write as if the second-hand evidence they cite from published reports on seventy-eight languages corroborates the results they worked out from their first-hand data for twenty languages, but that claim can scarcely be taken seriously. As they themselves point out, the published reports hardly ever specify the focal point of a given exotic

colour term, but rather list English words describing the total area covered by the term. Thus a four-term system might have a word glossed 'blue, green'; since Berlin and Kay believe that blue occurs as a *focal* colour only in systems of six or more terms, they count the word in question as meaning basically 'green' and claim the language as another instance of their 'Stage III' (black, white, red, green-or-yellow) of colour-term evolution – thus assuming what they set out to prove. A particularly flagrant example occurs in their analysis of the very restricted colour terminology of Homeric Greek, where the word *glaukos*, normally regarded as meaning 'gleaming, silvery' in the Homeric period and 'blue-green, grey' later, is said by Berlin and Kay to mean 'black', apparently for no better reason than that their theory demands a word for black and for some reason they have overlooked the existence of the standard Greek word for black, *melas* (despite the fact that this is actually by far the commonest single colour-word in the Homeric texts).[5] Clearly we must ignore the 'evidence' of the seventy-eight 'second-hand' languages and assess Berlin and Kay's theory exclusively on the data of the twenty languages which they examined in person.

But even their analysis of these languages contains many questionable points, for instance with respect to decisions (often crucial for their theory) as to whether some word is 'basic' in a given language or not. Frequently they seem to be led into error by ignorance. For instance, Berlin and Kay eliminate terms as 'non-basic' when they can be seen to be borrowed from other languages; but, while they are able to detect borrowings into various languages from English and Spanish, they appear not to realize that many terms which they list as basic for Vietnamese are borrowings from Chinese. If the Chinese borrowings were eliminated, Vietnamese would be left with words for black, white, red, brown, purple and grey, which would be a disaster for their theory. (Similarly, in discussing the language of Murray Island, New Guinea, which is one of the seventy-eight 'second-hand' languages, they eliminate many colour expressions on the ground that they are reduplications of nouns for objects having the colour in question – e.g. *bambam* 'orange, yellow', from *bam* 'turmeric'; but when it comes to *golegole*, 'black', which they need as a basic term, they dismiss as

'suspicious' the claim made by their published source that this derives similarly from *gole* 'cuttlefish' – Berlin and Kay perhaps do not know about the black ink secreted by cuttlefish.)

In other instances their decisions seem merely capricious. Thus, one of the traditional 'Five Colours' in Chinese thought, *ch'ing*, is commonly glossed 'green, blue, the colour of Nature'; Berlin and Kay list it as a basic term for the Cantonese dialect of Chinese and for Vietnamese and Korean (both of which borrowed the word from Chinese), in all three cases with the focal example in the same small area of the chart (a deep blue-green), but they count it as meaning 'blue' in Korean, 'green' in Cantonese and Vietnamese, and in Mandarin Chinese they ignore it altogether while including *lan* for 'blue', a word which etymologically referred to the indigo plant and is usually regarded as a subdivision of *ch'ing* (and which they explicitly omit for those reasons in their discussion of Cantonese).

Furthermore, Berlin and Kay seem not to appreciate the extent to which common traits in modern colour terminologies are influenced by the spread of a common technology with its range of pigments and dyestuffs, colour-coded electrical wires, traffic lights and the like. This effect is likely to have been particularly important in their research, since all but one of their twenty languages were studied through informants who lived in or near San Francisco. Noriko McNeill (1972) makes a relevant point here: she explains that the 'standard' eleven-term system which Berlin and Kay describe Japanese as possessing dates only from Japanese contact with the West, beginning in the 1860s, and that the traditional Japanese system of colour names has five terms whose foci are black, white, orange, turquoise, and yellow. This system is very awkward for Berlin and Kay's theory, but it is explained by the fact that the colours other than black and white correspond to natural dyes occurring in Japan.

Considerations of this sort go a long way also to explaining the ordering of the 'universal colours' described by Berlin and Kay. Thus, one puzzling feature of their ordering is the 'recessiveness' of blue: it is claimed to occur only in sixth place, after red, green, and yellow. At first sight this seems a quite surprising, unpredictable fact, if true, and thus a fact which counts heavily against the Whorf hypothesis; after all, blue is a primary colour, and one might suppose it merited a name as

much as red, green, and yellow. But how many blue things are there in the environment of a primitive culture? Sky and sea; but everyone knows their colour, so there is no point in discussing it. A few flowers perhaps, but flowers are of little practical importance; and the edible parts of plants, which do need to be discussed frequently, are never blue. Even in our own generation with its sophisticated chemical technology, blue is recognized by the manufacturers of commercial pigments as a difficult colour to create; it is small wonder that many primitive civilizations have got along without a special word for blue.

I have not yet dealt with the most striking of the facts presented by Berlin and Kay, namely that colour foci of diverse languages cluster in very limited areas of the colour chart. However, the explanation of this finding, which was provided by George Collier (1973), constitutes the most damning criticism of all.

The fact is that the variables of hue and tone are not the only variables relevant to colour. There is also the variable of 'saturation', which measures the extent to which a shade of a given hue and tone departs from the grey of the same tone. When we call a colour such as pillar-box red 'bright' or 'vivid' we normally mean not that it is light in tone but that it is highly saturated; 'old rose' would be an example of a low-saturation red. (In fact there is at least one other relevant variable besides hue, tone, and saturation, but we may ignore this.) Now the human eye is physically capable of perceiving greater saturation for some hue/tone combinations than others; a red of medium tone can be very saturated indeed, but even the 'brightest' light blue will not be too different from a light grey. Other things being equal, a language will obviously have names for the most vivid, noticeable colours rather than for colours in which high saturation is impossible. In other words, the continent which colonies divide between them is not a featureless steppe after all, but contains small areas of lush valley land alternating with wide, barren uplands. Quite naturally the first colonies will be founded in the best areas, subject to the constraint that no two colonies will be too close (it would be inefficient to have separate names for very similar shades of colour in a language which has few colour terms in total); only if there are many colonies will the middle slopes be occupied, and the highlands will always remain as tributary areas. Comparison of a chart of

attainable saturation at different hue/tone combinations with Berlin and Kay's chart of the distribution of focal colours shows the two to coincide almost perfectly.[6]

Taken in conjunction, these arguments seem to undermine Berlin and Kay's theory fairly completely. I have no doubt that the traditional, common-sense Descriptivist view of semantic variation is correct: where a fairly concrete domain of meaning contains no natural boundaries or specially salient features, nothing in our minds forces us to analyse it in one way rather than another, and languages will differ randomly in the way they categorize such domains. Cases such as the physics of space and time are very different. After all, most people recognize that the question 'How many colours are there?' is meaningless unless asked in the context of some particular principle for individuating colours, while we certainly do not regard the search for correct ideas about space and time as a meaningless activity, even though we may recognize that Mankind has not yet completed this search (and perhaps never will complete it).

So long as Sapir and Whorf claim only that our mother tongue provides an arbitrary but convenient set of pigeonholes for categorizing experience, on which we tend to rely whenever it appears to matter little what particular scheme of categorization we use, they are surely right. No doubt they are right, too, to argue that the decisions which we allow our language to pre-empt in this way sometimes matter more than we realize. But when they suggest that we are the helpless prisoners of the categorization scheme implied by our language, Sapir and Whorf underestimate the ability that individual men possess to break conceptual fetters which other men have forged.