EXPLORATORY ACCROSS-STIMULUS STUDIES IN EVENT-RELATED POTENTIALS

Malcolm Philip Young

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

1990

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/14740

EXPLORATORY ACROSS-STIMULUS STUDIES IN EVENT-RELATED POTENTIALS

Malcolm Philip Young

Psychological Laboratory,
University of St. Andrews,
St. Andrews, Fife,
Scotland, U.K.

Declaration for the degree of Ph.D.

I, Malcolm Philip Young, hereby certify that this thesis has been composed by myself, that it is a record of my own work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Signed Date 26th MARCH 1990

I was admitted to the Faculty of Science of the University of St. Andrews under Ordinance General No. 12 on 1st October 1987 and as a candidate for the degree of Ph.D. on 29th SEPTEMBER 1988

Signed Date 26th MARCH 1990

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the degree of Ph.D.

Signature Date 2 May 1990

UNRESTRICTED

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any <u>bona fide</u> library or research worker.

Signed Date 28th March 1990

"Never a man his making was made,
by living a life of ease" - Rudyard Kipling

"Absorb what is useful, reject what is useless,
and add what is specifically your own" - Bruce Lee

This work is dedicated to my wife, Debbie, and to my mother, Shirley.

Acknowledgements

I am grateful to Ivor Pleydell-Pearce, David Routh and Cheyne McCallum for initial inspiration, to A.D. Gordon and R.M. Cormack for statistical reassurance and to John Henderson and Phil Benson for help with some computer problems.

My thanks to Douglas Potter, Charles Pickles and Mike Doyle, my cohabitees in the Wellcome Brain Research Group office, for putting up with the music lover who occupied the office during the writing of this thesis.

I am grateful to Chris Wood, Greg McCarthy and David Perrett for useful discussions on topics connected with what follows and to Anne Perrett who perspicaciously aided the extirpation of profligate avocation of unnecessarily labyrinthine or obfuscatory parlance and who spotted the many split infinitives.

Most particularly, I am grateful to Mick Rugg. Mick carefully balanced rigorous argument and support, and provided the ideally challenging environment in which to learn how to do science.

I am grateful to the United Kingdom Science and Engineering Research Council for financial support.

I'm always amazed at how lucky I am to have Debbie's love and support and I'm sorry about the times that a busy internal life caused neglect of the external one.

CONTENTS

Part One

Part Two

FIGURES and TABLES

Chapter 2

Chapter 3

Experiment 1

Experiment 2

Experiment 3

Chapter 4

Experiment 4

Experiment 5

Chapter 5

Chapter 6

Chapter 7

ABSTRACT

Event-related potentials (ERPs) were evoked by visually presented words in a number of experimental paradigms. The quesion of which linguistic factors, if any, underlie differences between visual word ERPs was addressed. These studies identified 3 factors as predictors of ERP variance.

Studies of ERPs in language processing tasks are selectively reviewed, and methodological problems associated with ERPs evoked by non-identical stimuli are discussed. The importance of an understanding of the linguistic factors which underlie ERP differences is outlined, and a methodology for approaching this issue is set out. The statistical procedure necessary to address the question is developed and described in Chapter Two. This procedure was a quantitative modelling strategy, based on multidimensional scaling and PROCRUSTES rotation.

Five quantitative modelling studies were undertaken. These experiments involved two experimental tasks, a passive exposure task in which the subjects attended but did not respond to the stimuli (experiment 1) and a category membership decision task (experiments 2 to 5). Words drawn from two semantic categories were employed. ERPs were evoked by individual members of the category of colour names (experiments 1 to 3) and by members of the category of furniture terms (experiments 4 and 5). The results of these studies suggested that word length was the important factor in the early part of the post-stimulus epoch and that this factor was followed by semantic similarity. A late positivity was present in the decision task ERPs whose modulation was related to word frequency.

These results were validated by two conventionally analysed experiments which examined the relation between word length and repetition and that between word frequency and repetition.

It is concluded that three factors underlie ERP variance in the experimental paradigms employed. These factors are word length (physical extent was not dissociated from length in letters), word frequency and semantic similarity.

These results may inform issues of experimental control in future studies of ERPs and language processing, may suggest some reassessment of existing studies in which control was not effected for these factors and may have provided a method of wider utility in cognitive neuroscience. The results suggest that systematic information can be derived about the linguistic characteristics of individual words from single word ERPs.

PART ONE

Chapter One

Introduction

Introduction

The brain is an electrochemical machine. By a happy accident, some of the processes by which it operates produce electrical signals which can be recorded from the human scalp. This work takes advantage of this fact to investigate some aspects of the brain processes involved in visual word recognition.

The electroencephalogram and event-related potentials

The electrical signals emitted to the scalp by the brain are typically, in the case of brains without neurological abnormality, in the microvolt range. For this reason the first investigation of the electroencephalogram (EEG) lagged behind that of the observation of action potentials in peripheral nerve fibres, which could be investigated with relatively insensitive string galvanometers (Lindsley and Wicke 1974). An early success, however, was a study by Caton (1875), who observed both changes in electrical activity on sensory stimulation, and a continually changing background of activity in monkeys and rabbits. This was achieved with only a galvanometer and a magnifying glass for 'amplification' (Lindsley and Wicke 1974).

More detailed characterisation of changes in the electrical fields at the scalp, and their first recording in human subjects, awaited the development of suitable amplification equipment (Cooper, Osselton and Shaw 1974). Berger (1929) is generally accredited with priority, although his work did not become widely known until the publication by Adrian and Matthews (1934) of a series of replications of his observations.

The ’spontaneous’ EEG is the ’background’ sequence of electrical field changes which need not be directly related to sensory stimulation (Berger 1929; Lindsley and Wicke 1974). This takes the form of a series of transient voltage fluctuations whose frequency spectrum contains oscillations from very low frequency (with a time base in minutes (Cooper et al 1974)) to oscillations at tens of Hertz. Some information can be derived from analysis of the spontaneous EEG (Beaumont 1983), such as may be important for clinical use, or which may relate to gross behavioural states of the subject. But deflections of the spontaneous EEG are the result of the summation of signals from multiple processes in the brain. This mixing of signals often results in the occlusion of activity derived from particular processes of interest, and reduces the interpretability of the spontaneous EEG.

For example, suppose that a subject undertakes some prescribed task in which extraneous distraction is reduced to a minimum. His or her brain is still engaged in many processes which may be irrelevant to those in which the experimenter is interested. As well as responding to the task demands, while seated in a sound attenuated cubicle, at a table on which a T.V. monitor is placed, the subject has to maintain heart rate and breathing, a host of other vegetative functions, to regulate posture, to control blinking (see the method sections of Chapters 3 to 8) and so on. Many of these ’irrelevant’ processes will be reflected in the spontaneous EEG, and will, in consequence, mask part or all of the task-related processes which the experimenter is endeavouring to investigate.

Not all of these 'irrelevant' processes, however, will be related to the onset and offset of the stimuli on which the subject operates in the task. Only systems which are related to the processing of the stimuli will be time-locked to their appearance on the screen (Cooper et al 1974). It is possible to use this fact to segregate the scalp actvity derived from processes engaged by the task (the 'signal'), from processes not time-related to the onset of the stimuli (the 'noise').

In practice, this segregation is achieved by recording an epoch of EEG following the presentation of each of a class of stimuli. The length of this epoch is determined by the likely time course of task-related processing. These single trial EEG data are then averaged. Unrelated activity, which will be different in each sample, is diminished relative to activity which is constant (or relatively unchanged). This improvement in the signal to noise ratio is proportional to $n^{1/2}$, where n is the number of trials (assuming unlimited bandwidth) (Cooper et al 1974). The resulting 'event-related potential' (ERP) waveform is a partial record of the activities of those neural systems which were engaged in a time-locked fashion in stimulus processing.

The ERP waveform is a *partial* record because some systems involved in task-related processing need not give rise to scalp recordable activity. This could be for a variety of reasons (Donald 1979). Some neural populations might not influence potentials at the scalp because of their histological structure (i.e. 'closed field' structures (Wood 1987)); although both cerebral cortex and some medial temporal lobe (limbic) structures, such as the hippocampus and amygdala, are mainly composed of tissue which could give rise to electrical fields at the scalp (i.e. 'open field' structures) (Halgren and Smith 1987; Wood 1987; Wood, McCarthy, Squires, Vaughan, Woods and McCallum 1984). Alternatively, two or more generator populations may be oriented in the brain such that their electrical fields mutually interfere. In any case, it cannot be guaranteed that the ERP waveform is an exhaustive sample of those processes which mediated brain responses to the stimulus (Donald 1979).

The event-related potential is often characterised as a series of 'components'. These are manifest in voltage deflections either to the positive or to the negative at various latencies after stimulus onset. The naming conventions of these components are to label each with either an 'N' or 'P' depending on its relative polarity. The latency of the component, or its ordinal position in the sequence of deflections, is appended to signal its position in time. Thus, P300 or P3 is the positive deflection peaking at 300 msec post-stimulus, or the third positive deflection after stimulus presentation. In many cases, however, the actual peak latency may be secondary: P300 for example, varies in latency from about 300 msec to over 800 msec, depending on task difficulty and the subject population, without losing the designation 'P300' (Rugg, in press).

A particular voltage deflection at the scalp may reflect the influence of several distinct generator populations whose activities may overlap in time. Teasing these sources of variance out of the ERP records is a very difficult task, as is the identification of a particular component with a particular neurological structure, or structures (Kutas and Hillyard 1984; Wood et al 1984; Wood 1987). The conception of 'components' as the electrical resultants of neurologically distinct sources which summate to give rise to scalp recorded deflections (e.g. Naatanen 1982; Naatanen and Picton 1986) is therefore difficult to use (Gaillard 1988).

There are two alternative ways to conceptualise components. One is to define components as regions of variance in the waveform which are modulated by particular experimental manipulations or task demands (Desmedt 1981). This approach also has problems: it may be that descriptions of the effects of manipulations on components are logically circular when the component is itself defined in terms of the manipulation (Gaillard 1988). Another approach is to employ the convention that a component is directly manifest in a voltage deflection, or in a difference waveform between two conditions (Ritter, Simson and Vaughan 1983), unless some specific evidence for the presence of dissociable subcomponents of the deflection is present (Kutas and Hillyard 1984). I intend to use the term 'component', in this latter way.

Another organising principle which is often applied to event-related potentials is that of a dimension between 'exogenous' and 'endogenous' components of the waveform (Desmedt 1981; Donchin, Ritter and McCallum 1978; Rugg, in press).

The exogenous components are in general of shorter latency than endogenous components, and are modality specific. They can be conceptualised as obligatory, 'hardwired', consequences of suprathreshold stimulation (Desmedt 1981; Gaillard 1988), and they tend to be modulated by changes to the physical characteristics of the stimuli (Curry and McCallum 1982; Gaillard 1988). Some examples of exogenous components are the brainstem auditory evoked potentials (Jewett, Romano and Williston 1970), or the early components of the somatosensory ERP (Desmedt 1981). In some cases exogenous components are so insensitive to the behavioural state of the subject that they can be observed in comatose patients (Chiappa 1983).

In the visual modality, the modality with which all the studies presented here are concerned, the exogenous potentials are dependent on the type of stimulation. Flash evoked ERPs, for example, are highly variable between subjects and may evoke between 5 and 7 deflections in the first 250 msec post-stimulus (Starr, Sohmer and Celesia 1978). Visual ERPs evoked by patterned stimuli, such as words, however, evoke a negative deflection at about 70 msec, a positive one at about 95 msec, a larger negative deflection at approximately 150 msec (often [sic.] termed 'N1'), and a positive deflection which peaks at about 200 msec (Starr et al 1978; Cooper 1982; see Chapters 3 to 7). There is evidence that the mid-latency P200 is still exogenous in that it is modulated by physical differences between stimuli (Kutas and Hillyard 1980b; Chapter 7). As there is also evidence that there are endogenous components with shorter latencies than P200 (Harter, Aine and Schroeder 1982; Ritter, Simson and Vaughan 1983; Ritter et al 1984) there is evidently not a sharp temporal dividing line between the exogenous and endogenous visual potentials.

The endogenous components are thought not to be directly influenced by the physical parameters of stimuli, and their modulation is dependent on the characteristics of the cognitive processes engaged by a task (Donchin et al 1978; Gaillard 1988). Endogenous components are often considered modality aspecific, though this is not true in all cases (e.g. N200: Ritter et al 1984). Depending on variables such as task difficulty, the morphology and latency of endogenous components may vary widely (Gaillard 1988). In circumstances in which the underlying processor is still engaged, some endogenous components (e.g. N200 and P300) may be elicited even in the absence of a stimulus (Sutton, Tueting, Zubin and John 1967; Curry and McCallum 1982).

Exogenous components, being modulated by the physical properties of stimuli, are easily manipulated experimentally. Endogenous components are purported to be associated with particular cognitive processes and are less easily manipulated. Cognitive processes can only be indirectly manipulated by arranging task structure in such a way as to engage particular hypothetical processors. Because of this, there is considerable debate about the psychological significance of almost every endogenous component, and some confusion about the criteria which must be satisfied before similar deflections in different circumstances are imputed to the same processor (e.g. Coles and Donchin 1988; Desmedt 1981; Gaillard 1986; Gaillard 1988; Kutas and van Petten, in press; McCallum 1986; Ritter et al 1984; Verleger 1988).

Notwithstanding these difficulties, I will place an organization on the endogenous components as follows. The earliest (post-stimulus) endogenous components are those associated with selective attention. These are an attention-related modulation at the latency of the visual P1 (Hillyard and Munte 1984), and a family of negativities, variously described as Nd (Hansen and Hillyard 1980) or processing negativity (PN) (Alho, Tottola, Reinikainen, Sams and Naatanen 1987; Harter and Previc 1978; Okita, Wijers, Mulder and Mulder 1985), or Na (Ritter, Simson and Vaughan 1983) (Gaillard 1988). These components often overlap the latency range of the visual N1 and P200 components, and were initially described as an enhancement, in the attended channel (of an auditory selective attention paradigm), of N1 amplitude (Hillyard, Hink, Schwent and Picton 1973).

In ordinal position of occurence after stimulus onset, the next phenomena are endogenous negativities which have been described as 'mismatch negativity' and the N200 (Ritter et al 1984). Although it is possible that there are multiple N200-like phenomena in the latency range between the P200 and the P300 (Naatanen and Gaillard 1983; cf., Donchin et al 1978) the antecedent conditions necessary to evoke a negativity in this window are quite well understood. An 'N200' is elicited by infrequent changes randomly occurring in a train of stimuli. The component is inversely related in amplitude to infrequent stimulus probability, is modality specific in scalp distribution, is seen whether the infrequents are attended or ignored, and is often associated with a subsequent P300 ('P3a', which is a sharp, frontally-maximum positivity probably distinct from the generic P300 ('P3b', see below)) (Reviews are found in Donchin et al 1978; Naatanen 1982; Ritter et al 1984; Rugg, in press). The peak latency of N200 has been found to covary with reaction time (Renault and Lesevre 1979; Ritter, Simson, Vaughan and Friedman 1979). There is a consensus that N200 reflects some aspect of a process of discrimination, or of orienting the processing resources of the subject to infrequent stimuli (Ritter et al 1984).

A rather later endogenous negativity, peaking at 400 msec post-stimulus, was first recorded by Kutas and Hillyard to semantically anomalous sentence endings (Kutas and Hillyard 1980a). The N400 has provided a central motif in the domain of ERPs and language, and is discussed at length in the next section and in Chapter 7. Discussion in the present context, that of its relation to the other endogenous components of the ERP, is limited to the following remarks. It is presently much contended whether the N400 is a component in its own right, possibly one 'dedicated' to aspects of semantic processing, or whether it is the same phenomenon as the N200, discussed above. The greater peak latency of the N400 is explained, by proponents of the latter view, to be due to the greater difficulty of discrimination on the basis of semantic cues as compared to discriminations based on physical ones. The N400, however, is later than the N200, differs in scalp distribution (Kutas, van Petten and Besson 1988), is evoked by different task manipulations (and hence might be expected to be evoked by the engagement of different cognitive processes), and is not necessarily followed by a P300 component (Kutas and Hillyard 1983). The best arguments for the identity of N400 and N200 seem to be made on the basis of taxonomic parsimony.

By far the most extensively studied endogenous component of the ERP is the P300, or P3b (Coles and Donchin 1988; Desmedt and Debecker 1979; Donchin et al 1978; Duncan-Johnson and Donchin 1977; McCarthy and Donchin 1981; Rugg, in press; Sutton et al 1967; Verleger 1988). This is a modality aspecific late positivity, which is generally maximal over the midline parietal electrodes (cf., Johnstone 1989). Its peak latency is sometimes correlated with reaction time, particularly when the task modulates processes which may be imputed to stimulus evaluation rather than to response choice and execution (McCarthy and Donchin 1981). For this reason it has been exploited to investigate the possible locus of the effects of several factors which are known to affect reaction time (e.g. word frequency: Polich and Donchin 1988; Stroop interference: Tominaga, Kuda, Yogi, Miyara, Tomori 1989).

P300 has been invested with a large number of processing significances (Curry and McCallum 1982; Gaillard 1988). These include resolution of uncertainty, postdecisional closure, context updating, decision making, expectancy, surprise, stimulus evaluation and equivocation (Curry and McCallum 1982; Verleger 1988). At present, the precise cognitive operation, or operations, which are signalled by the occurrence of a P300 are still not agreed, indeed recent discussions of these issues have become rather intemperate (Coles and Donchin 1988; Verleger 1988). In contrast, the antecedent conditions for the evocation of a late positivity are, as with the N200, quite well understood. P300 is evoked by subjectively improbable stimuli. In general its amplitude is inversely related to the subjective probability of the stimuli, and directly related to their task relevance (Donchin et al 1978; Duncan-Johnson and Donchin 1977). In the case of task 'irrelevant' stimuli it is necessary that such stimuli engage the processing resources of the subject (Donchin et al 1978), perhaps by requiring some processing in order to reject them as targets (see Chapter 7).

In addition to the endogenous components discussed above, there are a number of slow potential shifts which occur prior to, or after, stimulus presentation or motor output. These components do not bear on the present work and will not therefore be discussed (but see for review Curry and McCallum 1982; Donchin et al 1978; Gaillard 1986; McCallum 1986).

Event-related potentials in the study of language

Research on the relation of ERPs to language processes has proceeded for almost 30 years and has produced a very large literature. One way to organise this domain is to describe it as being segregated into phases, which can be characterised by the aims of the research.

In general there have been two aims in ERP studies of language processing. The first has been the attempt to identify the **neuroanatomical locus** of language-related ERP effects. This has, in practice, been manifest in the search for hemispheric asymmetries in processing, and hence for ERP correlates of the different processing emphases of the two cerebral hemispheres (Galambos, Benson, Smith, Schulman-Galambos and Osier 1975; Kutas and van Petten, in press; Molfese 1983; Rugg, Kok, Barrett and Fischler 1986).

The second phase of research into ERPs and language has been to investigate the **constituent cognitive processes** of language tasks, particularly those derived from behavioural research. This approach proceeds without the need for a priori knowledge of the likely intracerebral generators of the ERP, and takes advantage of the good time resolution of ERP techniques and their (relatively) unobtrusive nature (Kutas and Hillyard 1984; Kutas and van Petten, in press; Rugg et al 1986). The basic assumption in this area is to assume that ERP differences are related to neural systems which have a functional significance in the cognitive processes of the subject (Kutas and Hillyard 1984). Reliable ERP differences are then interpreted as indicants of the differential processing of the classes of stimuli, or task conditions, from which the waveforms were evoked (Kutas and Hillyard 1984; Rugg 1987). The point in time at which two waveforms diverge is taken to give an upper bound estimate of the time-course of the putative mechanisms mediating their processing (Meyer, Osman, Irwin and Yantis 1988; Rugg 1987).

Several reviews of the ERP and language domain have dealt with the period in which the search for lateral asymmetries of processing provided the main focus of research. These reviews make depressing reading. Despite hundreds of papers, and more than a decade of effort, the evidence is (still) unconvincing that ERPs reflect the well-known asymmetries (Galambos et al 1975; Kutas and van Petten, in press; Molfese 1983; Rugg 1983b; Rugg et al 1986).

The clinical and behavioural evidence for hemispheric asymmetry is strong (Gazzaniga 1984). The absence of reliable hemispheric asymmetries in ERPs recorded in appropriate tasks, therefore, leads to a number of possible conclusions. Galambos et al (1975) argue that either the ERP technique is "virtually blind" to processing differences of this kind, or hemispheric differences are scarcely present in the paradigms on which their review was based. Latterly, converging evidence can be adduced in favour of both possibilities. Cerebral blood flow (rCBF) and positron emission tomographic (PET) studies indicate that many language functions are not as rigorously lateralised as might have been thought on the basis of the clinical and behavioural evidence: language processing can and does engage processors in both cerebral hemispheres (Ingvar 1983; Peterson, Fox, Posner, Mintun and Raichle 1988). Hence, the lack of asymmetrically distributed ERP components in language tasks may reflect the lack of differential engagement of the two hemispheres. On the other hand, depth recordings in monkeys and humans, and magnetoencephalographic (MEG) studies, both of which allow better spatial resolution than is possible for ERP data, suggest that, due to volume conduction, the signals from unilateral generator populations may in some circumstances be recorded from both hemiscalps (Arrezzo, Vaughan and Koss 1977; Arthur, Lewine, Schmidt, Oakley, George, Roeder, Hillyard, Aine, Kutas and Flynn 1989; Barrett, Blumhardt, Halliday, Halliday and Kriss 1976; Kutas and van Petten, in press).

In recent years the attempt to find the neuroanatomical locus of language related ERP effects, and in particular the signs of hemispheric specialisation, has declined in significance. The development of MEG techniques may reopen interest in the former issues, though such efforts are likely to be framed in the perspective of more recent results (e.g. Arthur et al 1989). The emphasis, in the last decade, has been on the elucidation of the cognitive mechanisms which may subserve linguistic processes and comprehension.

This approach has been successful in documenting a large number of correspondences between the ERP and cognitive domains, which exemplarise, in some cases, very precise covariation between the two areas. I do not propose to review the entire field here (but see for review: Hillyard and Kutas 1983; Kutas and Hillyard 1984a; Kutas and van Petten, in press; Molfese 1983; Rugg, Kok, Barrett and Fischler 1986), as many currently explored aspects only tangentially relate to the central focus of my present concerns. Some central themes in current approaches, however, are germane.

Kutas and Hillyard (1980a) recorded ERPs to each word of a series of seven word sentences. Words were presented one at a time at a slow rate (the interstimulus interval (ISI) was 1000 msec) and were silently read by the subjects. When the terminal word of the sentences was presented in a typeface with larger lettersize than that in which the earlier words had appeared, the ERP derived from these words contained a late positive complex of waves. These components were very similar in morphology to the P300 which is paradigmatically recorded in conditions which disconfirm the expectations of the subjects, or which are subjectively improbable (Donchin, Ritter and McCallum 1978; Duncan-Johnson and Donchin 1977; McCarthy and Donchin 1981).

In sharp contrast to the ERPs evoked by the physically deviant words, sentence endings which embodied semantic anomalies (e.g. "He spread the warm bread with socks") elicited a monophasic negativity (i.e. one not followed by P300) which peaked at about 400 msec. This N400 component was interpreted, on the basis of the fact that only the semantically anomalous endings evoked it, as a sign of the ""reprocessing" or "second look" that occurs when people seek to extract meaning from senseless sentences" (Kutas and Hillyard 1980a).

Kutas and Hillyard (1980b) employed a very similar design but factorially crossed semantic and physical deviance. The late positive waves and the N400 elicited by the physical and semantic manipulations were found to be additive, a finding which was interpreted as a dissociation of the two phenomena. In addition this study recorded from lateral electrodes and found that throughout the sentences there was a hemispheric difference in potential, which took the form of greater negativity over the right hemiscalp.

An attempt at controlling for physical differences between the items in each condition was made by Kutas and Hillyard (1980b). This involved ensuring that the congruous and incongruous sentence endings did not differ in terms of their average word length. No mention was made of control for other possible differences, and the word lists were not appended. Control for the differences between large and small typefaces was undertaken by presenting the word 'STATION' to the subject 80 times in each of the two typesizes. In the latter case, comparisons of the ERPs evoked by the large and small 'STATION' conditions with the large and small sentence completion items do not focus only on the intended physical differences between them: such comparisons are conflated with differences of repetition, which have since been discovered to modulate the ERP waveform (Chapters 6 and 7; Rugg 1985; Rugg 1987).

The antecedent conditions necessary for evoking the N400 were further elaborated in Kutas and Hillyard (1983). In this study the task was made more naturalistic by having subjects silently read variable length sentences which comprised connected prose passages dealing with a single topic. The words were presented at a faster rate than previously, though still more slowly than in self-paced silent reading ($640 < \text{ISI} < 760$ msec). ERPs were recorded from both sentence ending words and the intermediate words in all sentences.

The main feature of the design was to compare the ERP concomitants of semantic anomalies with grammatical anomalies. The grammatical errors were noun number errors (singular/plural), verb number errors, and verb tense errors, all of which were placed at intermediate sentence positions. The results showed that while an N400 component was elicited as before to both intermediate and sentence-ending semantic anomalies, the grammatical errors did not evoke this component (though a sceptic would note that the verb tense errors evoked a deflection which looks like a small N400 (see Kutas and Hillyard 1983: Figure 4 and Table 2)).

In addition to the analysis above, the recording of ERPs to all words in a large number of sentences provided the opportunity to examine the differences between 'content' and 'function' words. These word categories, also known as the 'open' and 'closed' classes, respectively comprise the noun, verb and adjective constituents of language, and the articles, prepositions and auxiliaries which embed the content items into ordered sentences. ERPs evoked by these two classes of vocabulary were characterised by differences in scalp distribution. In particular the 'open' class words were found to evoke a greater sustained positivity beginning at about 200 msec than the 'closed' class words. This difference was clearest at the frontal leads and possessed an asymmetric scalp distribution, so that content items evoked a greater positivity over the left scalp sites.

It is likely that this difference in ERPs evoked by the two word categories is contaminated by effects due to other factors. The function words in a series of connected prose passages, where each passage deals with a single topic, are likely to be repeated more often than are the content words they embed.

Hence ERPs to these items are likely to be influenced by the fact that they have been repeated (see Chapters 6 and 7; Rugg 1987). Similarly function words are in general of very high written word frequency, when compared to content words. These conflated factors, however, are probably not responsible for the effect reported by Kutas and Hillyard (1983), indeed the probability is that, if they do influence the waveforms, they tend to undermine the real differences between open and closed class ERPs. The comparison between the two classes of words is conflated with the comparison between unrepeated low frequency words (the content items) with repeated high frequency words (the function items). There is evidence (Chapter 7; Rugg, in press) that this combination of repetition and frequency differences will contribute to an effect in the opposite direction to that of Kutas and Hillyard's (1983) open/closed class effect. Unrepeated low frequency words evoke a sustained *negativity* compared to repeated high frequency items. It is possible, however, that the apparent anterior distribution of the open/closed class effect is due to the superimpostion of the centro-parietal effects of word frequency and repetition on an effect due to word class.

Kutas and Hillyard (1984) returned to a paradigm based on the presentation of unconnected sentences. In this study the amplitudes of the N400 components evoked by semantically deviant final words were found to be closely related to their 'Cloze' probability. The Cloze probability of each word was determined by recording the number of times it was given as the terminal word, when a large number of subjects provided a word to complete the sentence (Bloom and Fischler 1980; Kutas and Hillyard 1984) (e.g. Many subjects would complete the sentence "He mailed the letter without a ..." with the word "stamp", while very few would complete the sentence with "cheque": thus "stamp" would have a high, and "cheque" a low Cloze probability for this sentence).

The fact that N400 amplitude was found to covary with Cloze probability, with a correlation exceeding 0.7, forced a reevaluation of the processing significance of the component (Kutas and Hillyard 1984). In the previous studies, the antecedent condition necessary for the evocation of N400 was thought to be semantic incongruity (Kutas and Hillyard 1984). The close covariation of N400 amplitude with Cloze probability, however, was taken to suggest that it varied systematically and continuously with the extent to which a word was expected in the context of the preceding sentence. Further, by examining the relation of N400 to low probability words which were related to the 'best completions' (words with the highest Cloze probability) of highly constrained sentences, it was found that N400 was sensitive to the semantic relatedness of the terminal word to the expected one. For example, for the sentence "He liked lemon and sugar in his ..." the word "tea" is the best completion, and evoked a small N400 component. But the word "coffee", which is semantically related to "tea", although somewhat incongruous in this context, also evoked a small N400. Hence, rather than interpreting the N400 as a signal of contextual violation, the preferred interpretation of Kutas and Hillyard (1984) was that the component reflected the degree to which a word is semantically primed (see below) by the preceding context.

Bentin, McCarthy and Wood (1985) investigated the modulation of ERPs by semantic priming in a word-nonword classification (lexical decision) task. Subjects were presented with a continuous sequence of words and nonwords, at a fixed intertrial interval of 2500 msec. Subjects were required to respond with a button press made with their preferred hand on deciding that a presented item was a word, and with a button press with their other hand on detection of a nonword. The stimuli were 80 'primes', which immediately preceded 80 semantically related (by membership of the same category) 'targets', 80 'fillers' which were semantically unrelated to either the primes or targets, and 240 pronounceable nonwords, constructed by permuting letters from the word stimuli. Control was effected for word frequency, word length and for sequence effects.

The behavioural results, the reaction times and error rates, followed the pattern of previous behavioural studies: subjects responded more quickly and accurately to semantically primed words than to the unprimed primes and fillers, or to the nonwords. The ERPs evoked under these conditions showed a negative deflection which peaked at about 400 msec. This negativity was significantly greater for the unprimed conditions than for the primed 'target' waveforms. The difference waveform derived by substracting the target and prime traces was found to possess a centroparietal, slightly right hemisphere dominant distribution.

These features of the priming modulated negativity at 400 msec were compellingly similar to the negativity recorded by Kutas and Hillyard in the sentence presentation experiments discussed above. The nature of the modulation of this negativity was also consistent with the relationship between N400 and semantic priming proposed by Kutas and Hillyard (1984) (Bentin et al 1985; Kutas and van Petten, in press). On the interpretation that the modulation by priming of an endogenous negativity in the 400 msec latency range was a modulation of N400, the results of this study supported the view that the prior presentation of a single related word may be sufficient to contextually prime a subsequent item (Kutas and Hillyard 1988).

Boddy, in a series of studies, explored the action of semantic priming on the N1 and P200 components of the visual evoked response (Boddy 1981; Boddy and Weinberg 1981; Boddy 1986). These studies involved either a category membership decision task (e.g. Boddy 1981) or a word-nonword decision task (e.g. Boddy 1986). Effects of the priming manipulation on N1 to P200 peak to peak measurements were interpreted as early 'attainment of meaning' (prior to 130 msec, being the approximate latency of N1; Boddy 1981). The N1 amplitude differences recorded as a function of the relation of the target word to the preceding item, however, were very small (0.5-1.0 uV; Boddy 1981). It is quite surprising that such small differences were statistically significant given that the ERPs were based on only 6-10 trials (Boddy 1981) (and that they necessarily therefore contained substantial amounts of noise), and the analysis of variance was based on only 5 subjects (Boddy 1981). It is possible that the uncorrected repeated measures analysis of variance applied to these data (cf. Keselman and Rogan 1980) was somewhat admissive. Additionally, the P200 differences in a later experiment (Boddy 1986) seem, by examination, to be modulated by "an enhanced negativity [which] appeared to begin before P200 .." (Boddy 1986 p 304). Thus it is possible that some or all of the effects attributed by Boddy (1981; 1986) to the N1-P200 peak to peak measurement were in fact due to the temporal overlap of an early onsetting N400 with P200 (see Chapters 6 and 7).

A further principal interest of Boddy's experiments was to explore the effects of differing stimulus onset asynchronies (SOA) on the ERP in a priming paradigm. This approach was based on the recognition that most studies in the domain of ERPs and language present the items at rates many times slower than that in normal reading (Boddy 1986). The effect of lowering SOA to an 'ecologically valid' 200 msec was to enhance the difference between N400s recorded to primed and unprimed words (Boddy 1986). There have, however, been some criticisms of these studies, based on the problems which attend the measurement and identification of deflections to the second word in a paired presentation procedure, when the SOA is so short that the ERP to the second word overlaps with that of the first (e.g. Kutas and van Petten, in press).

Rugg (1985; 1987) explored the relation between the modulation of the ERP by semantic priming and that by repetition. Notwithstanding the conceptual similarity between priming an item by a semantic associate and priming by the prior presentation of the item itself, the effects of these two manipulations on performance suggest a dissociation. While the facilitatory effects of semantic priming are short lived (on the order of a few seconds; Henderson, Wallis and Knight 1984), the effects of identity priming (repetition) can be demonstrated after long periods intervening between two presentations (on the order of several days; Scarborough, Cortese and Scarborough 1977). Moreover the effects on performance of priming and repetition have been reported to be additive (Den Heyer, Goring and Dannenbring 1985), which on some accounts (Sternberg 1969) is construed as evidence for a dissociation of the processing loci of the effects.

The ERPs reported by Rugg (1987) show the familiar pattern of the reduction of an endogenous negativity which peaked at about 400 msec poststimulus (N400) by priming the items with a semantic associate. The modulation of the ERP by repetition, however, took the form of a relatively much larger effect on this region of the waveform, so that repeated items evoked traces which were substantially more positive going than unrepeated items in a window from 250 msec to 900 msec. Additionally, and in contrast to the effects of semantic priming, there was a relative negativity present in the repeated traces which coincided in time with the peak of P200 (Rugg 1987). These differences of magnitude in the later components, together with the fact that the priming and repetition manipulations gave rise to different topographic distributions of the potential field over the head, were interpreted as signalling that semantic priming and repetition engage at least partially different processors. In subsequent studies, where priming and repetition have been factorially crossed (e.g. Kutas, van Petten and Besson 1989), the pattern of interaction between the two factors appears supportive of this view, although it is probable that the two manipulations share subcomponents, such as the diminution of a 'default' N400 in unprimed and unrepeated items (Rugg, Furda and Lorist 1988), (see the discussion section of Chapter 7 for a detailed treatment of this issue).

Language related ERP effects, such as those noted above, have been manifest as the modulation of a small number of ERP components rather than the finding of a plethora of new, 'dedicated', language-related components. Some components may be particularly engaged by experimental manipulations of a linguistic nature (e.g. N400: Kutas and Hillyard 1988), but the language-dedicated nature of even these phenomena are presently contended, both on the grounds that the N400 may be a species of N200 (Herning, Speer and Jones 1987; Kutas and Hillyard 1980c; Polich, Vanasse and Donchin 1981; Pritchard 1981; Ritter, Ford, Gaillard, Harter, Kutas, Naatanen, Polich, Renault and Rohrbaugh 1984), and on the grounds that 'priming' experiments with face and picture stimuli, which are not ostensibly linguistic, appear to evoke an N400 (Barrett, Rugg and Perrett 1987; Rugg and Barrett 1987).

This most recent phase of research, with its emphasis on task related aspects of processing, has succeeded to the extent that it is sometimes supposed that ERP differences may relate almost exclusively to the constituent cognitive processes of an experimental task, and not to the intrinsic characteristics of what is being processed.

There are very good historical reasons for the relative emphasis on task related aspects and de-emphasis on the 'content' of processing. These will be the focus of attention in the next section, but for the present it should be noted that this is not an entirely benign assumption. In the extreme it might tempt the unwary to forego careful control in list presentation experiments. Indeed, several important papers in which list items were not counterbalanced between conditions make no mention of factors for which control was effected in stimulus choice (e.g. Kutas and Hillyard 1980; Kutas and Hillyard 1983; Kutas and Hillyard 1984; Rugg 1984; Rugg 1985; but compare: Bentin, McCarthy and Wood 1985; Kutas and Hillyard 1988; Kutas, van Petten and Besson 1988; Rugg 1987; Rugg, in press). This leaves open the possibility that some effects have been wrongly imputed to experimental manipulations, when they may be due to badly controlled and unidentified covariate factors, or to the interaction of unidentified covariates with aspects of task demand. From this perspective it is of paramount importance to generate an insight into those intrinsic linguistic factors which are important in giving rise to ERP variance (Fischler 1987; Kutas 1987; Kutas and Hillyard 1983; Rugg, Kok, Barrett and Fischler 1986) so that, among other things, control may be effected for them when they are not the subjects of study.

Previous attempts at conducting across-stimulus studies

The focus of a third tradition of ERP language research was to identify neither the neuroanatomical locus nor the putative cognitive processes engaged by a task. It was, instead, to find ERP correlates of stimulus 'information content' (Chapman, Bragdon, Chapman and McCrary 1977; Chapman, McCrary, Chapman and Bragdon 1978; John et al 1967; Thatcher and John 1976; Purves, Low and Baker 1979). The difficult methodological challenges with which this approach was faced, which it did not succeed in overcoming, are addressed in this section.

Begleiter and Platz (1969) recorded ERPs in three experimental conditions. These conditions comprised repeated exposure to a list of (two) taboo words, a list of (two) neutral words and blank fields. Signals were recorded from one electrode and a simple peak to peak measurement analysis undertaken. This revealed that the N1 to P200 amplitude difference was larger for taboo words than for either neutral words or blank fields, which did not differ. The P200 to N400 difference was also larger when evoked by the taboo words than by the neutral words, though for this measurement the neutral words gave a larger amplitude difference than did the blank fields. Begleiter and Platz (1969) concluded that the connotative meaning of the stimulus was reflected in the ERP.

This study provides both a beginning to the tradition of across-stimulus studies which I am going to examine, and a very clear example of the interpretative problems associated with such experiments. Obviously the stimuli in the two word lists (of two items each) could not be identical. From the analytic strategy used, it can only be taken that ERPs evoked by different word lists differ; which of itself is not a finding of great profundity. Before concluding that ERP differences between lists were the result of **meaning** differences, it would be necessary to ensure that no other factor could be invoked to explain the presence of reliable differences. But taboo and neutral words might differ in a number of other ways. For example, taboo words are often short and structurally simple (orthographically or phonologically) in accord with their role in language as expletives. Taboo words are in general very low in written frequency (see Kucera and Francis 1967). In the latter case, the taboo words chosen by Begleiter and Platz (1969) had an average occurrence in the Kucera and Francis (1967) corpus of 3, and the neutral words of 41. Subsequent work (see Chapter 7; Rugg, in press; Smith and Halgren 1987) is quite consistent with the conclusion that both an enhanced P200 component and an enlarged N400 might be the result of systematic differences in word frequency between Begleiter and Platz's (1969) lists of taboo and neutral words.

Chapman and colleagues (Chapman et al 1977; Chapman et al 1978; Chapman 1980) attempted to explore the relation between ERP waveforms and Osgood's 'semantic differential' (Osgood 1971; Osgood, May and Miron 1975; Osgood, Suci and Tannenbaum 1957). The semantic differential was derived by the application of linear factor analysis to a large number of subjects' associations of bipolar adjectives (scales) with selected nouns (concepts) (Fillenbaum and Rapoport 1971; Rapoport and Fillenbaum 1972; Rosenberg and Sedlak 1972). The resultant orthogonal factors defined a 'semantic space', and enabled the assignment of a set of coordinates to each concept, which determined its position in this space. Three dimensions:

Potency (strong-weak, heavy-light, hard-soft..), Activity (fast-slow, active-passive, excitable-calm..) and Evaluation (good-bad, pleasant-unpleasant, positive-negative..) were claimed to emerge from a wide variety of candidate nouns and scales, and this three-dimensional structure was claimed to "remain valid for more than 23 language/culture groups" (Chapman et al 1977).

Chapman et al (1977) constructed two lists of 120 words. Each list consisted of six sets of words, each of which were derived from poles of the semantic differential. That is, the words were chosen on the basis that they scored high or low on one of the dimensions of the differential, and were neutral on the other two. For example, 'CARE' was a member of one of the 'Evaluation+' (E+) sets, and this word scored 1.9 on evaluation, 0.6 on potency and 0 on activity (Chapman et al 1977).

ERPs were evoked by exposure to these words, which were presented in random sequences. The subjects were required to pronounce each word aloud "toward the end of the 2.5 second interval between each word" (Chapman et al 1977 p40). Each subject was presented with 12 or 18 runs of two lists. The data analysis process concentrated on the ERPs evoked by the E+ and E- lists by developing 'templates' for the two conditions on the basis of three subjects. These templates were then applied to ERPs from all subjects (including the three used to develop it) with the aim of classifying the responses into two groups. It was found that 69% of the responses of the nine subjects who did not contribute to the template could be classified by this procedure into the correct groups. This was interpreted as support for "the conclusion that differences on the evaluative dimension of visually presented words result in detectable and statistically reliable differences in brain responses", and moved the authors to claim that "rather detailed and specific covariation between linguistic and EP domains has been found" (Chapman et al 1977).

Chapman et al (1978) sought to replicate the findings of the 1977 report. The details of the design of the experiment, the electrode placements, the task and the word lists were as Chapman et al (1977), but the data analysis strategy was intended to be more defensible than the correlational approach of the earlier study, which has subsequently attracted methodological criticism (e.g. Molfese 1983). Ten subjects' data were analysed in the following way: each subject's data were standardised in a z-transformation by subtraction of the mean and division by the standard deviation. These standardised data were then submitted to principal components analysis (PCA). The PCA generated 12 orthogonal factors which together accounted for 94% of the total variance. These orthogonal factors were then employed as the input variables for a series of discriminant analyses.

All the computed discriminant functions classified the ERPs into the appropriate categories at well above chance levels, in both unidimensional (e.g. ERPs to E+ vs E- words) and multidimensional (i.e. using all six possible categories) analyses (Chapman et al 1978; Molfese 1983). The canonical variables, those that 'told the difference' best between the classes, were not reported. The stability of the solutions was tested by means of jacknife cross-validation, and by employing the discriminant functions computed for one list of 120 words on the ERPs derived from the other (Chapman et al 1978). The median success rate across the analyses based on the list used to develop the functions, the 'other list' and the jacknifed analysis, expressed as a percentage of correct classifications, was 100% (Chapman et al 1978; Table 1). These surprisingly orderly results "suggest[ed] that internal representations of meaning can be assessed by analysing brain responses" (Chapman et al 1978).

There are a number of problems which deserve mention concerning this often cited (e.g. Hillyard and Kutas 1983; Kutas and Hillyard 1984; Molfese 1983; Rugg et al 1986) 'demonstration' that the ERP reflects Osgood's classification system. The methodological problems associated with the correlational analysis of Chapman et al (1977) have already been treated by Molfese (1983) and I will not reiterate them here. Instead I intend to focus on the general design of these studies and on the analytic strategy employed by Chapman et al (1978).

The first problem is with the naming task employed to ensure that each word was perceived (Chapman et al 1977; 1978). A naming task which requires the subject to pronounce the words within 2.5 seconds of stimulus onset, in a task situation in which the interstimulus interval is 3.5 seconds, risks contaminating the ERP data with signals of extracranial origin. It has been shown that electromyogram activity, the glossokinetic potential (derived from movement of the tongue) and respiratory potentials may precede the onset of speech by more than a second (Grozinger, Kornhuber, Szirtes and Westphal 1980; Kutas and van Petten, in press; Picton and Stuss 1984; Szirtes and Vaughan 1977). The epoch recorded in these studies was 510 msec (Chapman et al 1977; 1978), which places the later components of the evoked potential quite close to the probable onset of speech related activity, assuming that subjects delayed their pronunciation of each word until 2 seconds after stimulus onset. Whether they did, in fact, delay their response by this amount is unclear, particularly since Chapman et al (1978) remark only that "the words were spoken after the 510-msec EP interval" (p203). This procedure would be unlikely to be acceptable according to the methodological standards applied at the present.

The next issue concerns the fact that part of the data analysis strategy employed by Chapman et al (1978) was the derivation by principal components analysis of orthogonal factors which were then used as input variables for the discriminant function analyses. It has since been discovered, by extensive testing with data whose parameters were known, that PCA can misallocate variance between components (Wood and McCarthy 1984). Where PCA is the first part of a PCA-ANOVA process, this misallocation can be demonstrated to lead to large increases in the probability of type I error (incorrect rejection of the null hypothesis) (Wood and McCarthy 1984). Although the misallocation of variance between components, when expressed as a percentage, may be relatively small (8-10%; Wood and McCarthy 1984), misallocation may have dramatic results on the subsequent assessment of treatment effects (70-80% inflation of type I error; Wood and McCarthy 1984). It has also been suggested that ANOVAs on principal component scores may be biased toward exaggerating the statistical significance of treatment effects (i.e. differences between conditions) because variance due to experimental treatments is used to define the dimensions of measurement (Hunt 1979; Wastell 1981; Wood and McCarthy 1984).

No such simulation tests of the veracity of a PCA-Discriminant analysis strategy have been undertaken, partly because this procedure has not been widely used. The effects of the dispositions of these techniques to misrepresent the data cannot, therefore, be known with certainty. It is possible, however, that the first part of the statistical process chosen by Chapman et al (1978) may represent the data in such a way as to enhance artificially the apparent reliability of differences between conditions (Hunt 1979; Wastell 1981). The finding of such surprisingly successful classification of the principal components, at a median of 100% correct classification, may therefore point up a deficiency with the analytic procedures, rather than a remarkably noise-free correspondence between ERP data and Osgood's semantic differential.

A third issue, which is perhaps the most serious problem, concerns the extent to which these studies controlled for sources of extraneous variance. The basis of Chapman et al's (1977; 1978) approach to experimental control is captured by the following excerpts.

"The physical parameters of the stimuli (various letters which comprise the words) will vary, of course, from one word to the next and their effects will tend toward the same average EP for the various groups of words. In short, while the background EEG is being averaged to obtain EP's, the physical characteristics of the word are being averaged to control for their effects and the meanings of the words are being averaged to provide a common core of semantic meaning" (Chapman et al 1977, p38), "the use of a fairly large number of words makes it likely that the variation in stimulus attributes are distributed alike to the semantic classes" (Chapman et al 1977, p40), and "In order to control commonly confounding variables....the semantic classes were represented by a relatively large number of different words" (Chapman et al 1978, p196). Thus the expectation is that, by including 20 different words as the constituents of each 'meaning' set, any differences between conditions which are not due to differences of position in Osgood's semantic differential will be averaged out.

This is a forlorn hope, as evidenced by the effort required to construct only two sets of words according to differences on one dimension (e.g. word frequency, see Chapter 7) while ensuring there are no differences according to another dimension (e.g. word length. See Chapter 6 for the obverse case). Notwithstanding the above excerpts, Chapman et al (1977; 1978) attempted to control for physical differences between the sets of words by examining the average word length and the distribution of letters in each of the word sets. The distributions of word length and of letters were similar for the semantic classes. The authors, however, note the possiblity that there may exist problems with other covariate confounds "it is difficult to attribute any obtained differences to anything other than semantic processing....[But] This conclusion will be reasonable provided that there are no stimulus attributes which are correlated with the semantic categories" Chapman et al 1977, p39, my insertion).

In none of the studies of the relation between ERPs and the semantic differential by Chapman's group are the word lists provided so that subsequent researchers may examine them: it is not therefore possible to identify with certainty any possible covariate confounds in the six sets. A sample of four items from each set (of 20 items), however, was included in the Chapman et al (1978) report. Unfortunately, a gross word frequency difference exists between the list samples, which vary between an average of 281.5 occurrences, for the 'P-' list, and 16 occurrences, for the 'E-' list, in the Kucera and Francis (1967) corpus.

This reflects another problem with the analytic strategy chosen by Chapman et al (1978). Discriminant analysis only tests the success with which a pre-chosen factor can segregate the data into appropriate classes, and does not give the opportunity for other factors, with which the chosen factor may covary, to 'compete'. The Osgood model may fit the data tolerably well, but it remains open whether classes derived, for example, from word frequency would have segregated the data in a manner which would have resulted in greater variance-explained (i.e. whether frequency would have been a 'better fit' to the data). It is not disputed that meaning differences **might** contribute to the observed differences between ERP waveforms, but I do dispute that this can be known with confidence, since no attempt has been made to assess the degree to which the results are explainable by other possible sources of variance. As noted above in relation to Begleiter and Platz (1969), a central problem with the interpretation of these studies, which evoked language ERPs by exposure to nonidentical stimulus lists, is the possibility that ERP differences are due to unidentified covariate confounds.

Finally, one can also question whether the semantic differential is an appropriate conceptualisation of **meaning**, whether connotational or otherwise (Fillenbaum and Rapoport 1971). I do not feel that the **meaning** of 'BATH', is adequately captured by the characterisation that it is very positively evaluated, quite potent and is rather inactive (see Chapman et al

1977, Figure 1). It is possible that the differential captures some underlying affective regularities in the concepts that we apply, but it seems unlikely that so crude a characterisaton of meaning would be of use in cognition. Indeed, there is little evidence that the semantic differential is psychologically real in the way that other means of placing a spatial organization on the semantic domain may be. For instance, proximity of items in the space defined by the differential has not been demonstrated (to my knowledge) to predict either efficacy of semantic priming or degrees of Stroop interference in reaction time tasks. By contrast, the 'semantic distances', derived from semantic similarity judgements, which are the basis of multidimensional scaling representations of concept interrelations (Fillenbaum and Rapoport 1971; Shepard 1980; see Chapters 3, 4 and 5) predict both (Young and Routh, unpublished observations). There have also been problems associated with replicating the semantic differential, particularly when the data are derived from 'naturalistic' judgements (Rosenberg and Sedlak 1972; Nerlove and Romney 1972).

To an extent, some of these difficulties with the interpretation of experiments based on the presentation of nonidentical stimuli were apprehended long ago (Johnstone and Chesney 1974; Brown, Lehmann and Marsh 1980; Brown, Marsh and Smith 1976; Roemer and Teyler 1977). An attempted improvement on this paradigm was to employ homophones or homographs (words that sound or look alike) in conditions of differing contextual constraint. Unfortunately, the problem is not abolished by this device, indeed it introduces another possible source of conflation: that between semantic and syntactic differences between the comparison items (Molfese 1983).

Brown, Lehmann and Marsh (1980) provides an example of the fact that the interpretative difficulties associated with across-stimulus studies may extend to identical stimuli which are interpreted differently. This study was part of a series of experiments in which careful attention had been payed to control for physical differences between comparison items (Brown, Marsh and Smith 1976), for differences in the preceding contexts (Brown et al 1976; 1979), for the degree to which subjects were, and were not, able to anticipate the intended meaning of the word (Marsh and Brown 1977), and to the consistency across subjects of noun/verb differences (Molfese 1983). In experiment one of Brown et al (1980), 7 subjects heard blocked presentations of the phrases: "A pretty rose" and "the boatman rows", both of which contained the homophone /roz/. There were consistent differences between the noun and verb interpretations of this acoustically identical stimulus, as determined by PCA and ANOVA on the factor scores (see remarks above regarding Chapman et al 1978).

Plainly, as discussed by Molfese (1983), the two meanings of the homophone have different syntactic roles. It is open to question, therefore, whether ERP differences are due to differences in meaning per se, or to less specific noun-verb differences. This objection could have been met by using two of the noun interpretations of the homophone /roz/, such as "a pretty rose" and "chairs arranged in rows". As it is, these syntactic and semantic factors are conflated.

Another problem is that the two syntactic roles of the homophone may themselves have different frequencies of occurrence. Any ERP differences may therefore relate to differences in word frequency between the alternative interpretations. In the case of 'rose' versus 'rows', there is indeed a word frequency difference as the former was recorded 86 times, and the latter only 16 times in the Kucera and Francis (1967) corpus. This frequency difference is further complicated by the fact that both 'rose' and 'rows' have several interpretations. The problem that individual senses of a word may have different word frequencies can be ruled out only by careful choice of a homophone whose interpretations have similar occurrences in English. A third problem, connected to the problem of word frequency, is that in this case orthographic and letter frequency variables are conflated with differences of meaning. Although the /roz/ is a homophone, the two interpretations differ in their constituent letters. As there is some evidence that orthographic factors may intrude into phonological processing (Rugg and Barrett 1987; Seidenberg, Waters, Barnes and Tanenhaus 1984), orthographic and letter level codes can not be assumed to have no influence. No doubt some or all of these extraneous factors could be controlled, but at present the results scarcely compel the conclusions claimed.

In the visual modality, Johnstone and Chesney (1974) presented an identical symbol in two contexts. The symbol was ambiguous in that it could be interpreted either as a '13' or as a 'B'. The two contexts were arranged by embedding the ambiguous symbol in a stream either of other numbers (e.g. 11, 15) or other capital letters (e.g. A, H). The stimuli were then multiply repeated and an ERP averaged for presentations of the same stimulus in the two contexts. Subjects were not aware that the '13' and the 'B' in the stimulus sequences had in fact been an identical symbol. These differences were interpreted as evidence for "neural activity correlated with the meaning of the stimulus" (Johnstone and Chesney 1974).

If, however, the symbol was interpreted in the two contexts as being a quite different stimulus, and it was processed accordingly, it is not the case that all confounds save meaning-related differences have been controlled. As above, only physical differences have been controlled. For example, '13' occurs only 47 times in the Kucera and Francis (1967) corpus, while 'B' occurs so many times that only 'BE', 'BY' and 'BUT' account for 16,063 occurrences in the same corpus. There is therefore the possibility that the results of Johnstone and Chesney (1974) are due (again) to a gross frequency difference between the interpretations of the stimulus, and, in consequence, the conclusion that the ERP differences are due to the authors' preferred interpretation, of differences in meaning, is far from compelling.

The central point is that the conclusion, in the case of linguistic material where physical differences have been controlled, that any reliable differences between ERP waveforms **must** be due to differences in meaning is patently inadequate. This argument is made in the form of argument by elimination, yet not all the differences between words, save meaning, are eliminated when control is effected for their physical differences. The obvious counterexample is word frequency, though other dimensions such as imageability (see Rugg et al 1986, p 293) would suffice to refute this approach.

The basic problem: Multidimensionality

The issue which the above critique highlights, and which no across-stimulus study in the area of ERP and language has hitherto adequately addressed, is that of words' **multidimensionality.** Words differ from each other along a large number of dimensions. Rubin (1980), for example, identifies approximately 50 dimensions, of which the following is a representative sample: letter frequency, word frequency, phonology, orthographic regularity, word length, concreteness, bizarreness, imageability, meaningfulness, typicality, familiarity, age of acquisition, ease of predication and pleasantness.

In the circumstance that an **intrinsic** property of linguistic items is the focus of an ERP (or RT) study, or that, for other reasons, the items that will constitute the experimental lists are not counterbalanced between conditions, it is paramount that an insight into the intrinsic properties of words which are likely to interact with task demand is available (Fischler 1987; Kutas 1987; Whaley 1980). Without such insight, any manipulation which involves nonidentical stimuli (and even some manipulations which employ identical stimuli, as noted above) may carry with it uncontrolled factors which make interpretation of treatment differences very difficult indeed.

The basic problem (and a prescription) is articulated by Fischler (1987 p 380): "Both materials and subjects ... [can be] instances of 'classification' variables, where the factor is inherent in the object, and we cannot randomly assign subjects, say, to the bilingual group, or words, say, to the high frequency condition. Statistically, most of us are guilty of treating such manipulations as experimental ones and proceeding with our analyses of variance, **when something like regression analysis is called for.** (My enbolding) ... The more serious problem is that ... observed differences in performance or ERPs could be due to some factor other then the nominally manipulated one. Especially for naturally occurring language materials, there are frequent opportunities for such confounds."

One way to conceptualise my approach to the relation of ERPs to language, is to see it as an attempt to take seriously the suggestion of Fischler (1987) that "something like regression is called for". Specifically, the central focus of this work is to address the question: **What intrinsic features of linguistic material interact with task demand in such a way as to give rise to ERP variance?**

In the foregoing sections I have tried to indicate some problems with approaches to the electrophysiology of language, past and present, which point up the importance of gaining an insight into this question. In practical terms, I believe that there are two possible approaches to it.

The first possible approach is to try to manipulate a small number of factors while controlling, by item selection, for all the other factors. This procedure would be close to the methodological conventions in this area, in that conventional experimental designs using list presentation and ANOVA could be employed. Nonetheless, there are serious problems with this approach. The largest number of factors which could be interpretably crossed in a factorially designed ERP experiment is about three (see for example, Halgren and Smith 1987). Because the number of possible factors is large (Rubin 1980), this would entail a large number of experiments. Even assuming that this could be done, there would be very difficult problems to surmount in virtue of the covariant nature of many of the dimensions, which would make item selection for many dimensions virtually impossible.

Additionally, the ERP technique requires that, for this type of experimental design, a given dimension segments the linguistic domain into quite large groups of words. This is both in order to construct lists of sufficient length that an ERP can be averaged from presentations of their members, and in order to select items to control for other dimensions. Not all dimensions do this. For example semantic similarity judgements, which are known to be 'psychologically real' (see the preceding section) in the sense that they predict behaviour in other tasks, are private to the semantic category to whose members they relate. They are meaningless outside the category, and hence cannot segment language material into large enough groups of words so that control can be effected for other dimensions, and an ERP evoked. For this reason there may be important dimensions of language which are not explorable by means of item selection, list presentation and ANOVA.

The second possible approach is analogous to that taken by psycholinguistic researchers when faced by the identical problem in the behavioural domain. By comparison with work in the domain of language related ERP research, it is a more unconventional and radical solution than that rejected above. Whaley (1978), for example, addressed the issue of which linguistic factors influenced reaction time (RT) in word-nonword classification tasks. This study was designed to find whether a manageable number of factors had an effect on RT, after the finding that RT was affected by word length, word frequency, and the frequency of the first and last letters of the word (respectively: Forster and Chambers 1973; Stanners, Jastrzembski and Westbrook 1975; Stanners, Forbach and Headley 1971). If a relatively small number of factors accounted for most of the variance in RT, then the large number of other possible factors could be safely disregarded in item selection. If, however, a large number of factors modulated RT, then item selection to control for those factors not of interest would be very difficult, and there would be some question of the suitability of this task in lexical research (Whaley 1978).

The approach to this problem was to generate a numerical model of the variance in response to a number of words. In this case the numerical model simply consisted of the mean RT of 32 subjects to 100 words. This model was then compared to numerical models of the same words according to the various dimensions in which they differed. Thus the RT variable was related to variables for word length, word frequency, letter frequency and so on. This comparison was achieved by multiple regression techniques. The outcome of this analysis was that, of the 16 linguistic factors tested, only three were required to account for most of the RT variance; these factors were word frequency, word length and 'richness of meaning' (Whaley 1978). Studies with a very similar rationale (Gernsbacher, unpublished and described in Gernsbacher 1984; Rubin 1980) have been undertaken with similar results.

This quantitative modelling approach, which psycholinguists have applied successfully to the behavioural domain, is the prototype on which I have chosen to model my approach to the problem of identifying linguistic factors in the ERP domain. Although the details of the actual technique are necessarily rather different (see Chapter 2) the underlying rationale is identical. The process consists of constructing a quantitative model of variance between conditions, and of comparing this model with models derived from the linguistic factors which the words embody.

A quantitative approach to across-stimulus studies

Part One describes a series of experiments which differ in two respects from most other research using event-related potentials.

The first difference is that the ERPs are not evoked by exposing a subject to a list of items of some experimentally manipulated class. These experiments, by contrast, involve ERPs evoked by exposure to repeated individual words.

The second difference follows from the fact that individual words differ one from another along a large number of dimensions (as discussed above; Rubin 1980; Whaley 1980). It is insufficient to employ an omnibus test, such as analysis of variance (ANOVA), as the sole analytical technique in this instance, because the particular factors responsible for ERP variance would not be identifiable. Hence a statistical rationale which eventuates, not in an essentially qualitative and binary decision as to whether conditions differ, but in a **quantitative model** of variance between conditions is employed (see Fischler 1987).

Another way to express this fundamental idea is to note that the statistical question addressed by the approach in Part One is not whether it is the case

that conditions differ, but is concerned with **how and in what way** they differ. By comparing the structure of differences of ERPs evoked by different words with the ways in which the words would differ according to a number of linguistic factors, it is hoped that the factors which 'drive' ERP variance will be identifiable. In this way the influence of an extensive set of variables on the ERP variance model may be analysed in terms of variance explained (Whaley 1980).

In a conventional list-presentation experiment, extraneous factors could be controlled in favour of the experimental manipulations by careful selection of the items which constitute the lists. Obviously, no comparable control procedure can be achieved for ERPs evoked by individual words. A quantitative modelling approach, instead, effects a **statistical control** in an experiment: irrelevant factors will explain little variance. This kind of approach, as indicated above, has been used successfully and extensively elsewhere, in the social sciences, in psychology and in neurobiology (Gernsbacher 1984; Joreskog and Sorbom 1984; McDonald 1988; Rubin 1980; Whaley 1980; Yamane et al 1989).

A two part thesis

No previously documented attempts have been made to apply this type of quantitative modelling approach to single word ERPs. Because of this, there exists a problem of two unknowns: the research will proceed by applying a novel, and therefore unknown, methodology in a novel and unknown area.

To address this difficulty, the work is divided into two parts. In Part One I develop a new quantitative modelling technique, specifically geared to the problems associated with ERP data and designed to answer the question identified above. I then apply it to empirical data in a number of experiments. In Part Two I test the veracity of the results of Part One by means of conventional, factorially designed, experiments, analysed by statistical methods whose propriety is universally recognised in this field. In this way, Part Two will serve to corroborate, or identify problems with, the technique and results derived in Part One.

By examining a restricted domain of language which is well characterised in terms of its psychological properties, it is hoped that these exploratory studies will constitute an extension of the 'method of single cases' (Clark 1973) to the cognitive psychophysiology of language.

Chapter Two

Statistical methods

Statistical methods

In Chapter 1, the central question with which Part One is concerned was defined: What are the linguistic factors which interact with task demand in such a way as to give rise to ERP variance? The general statistical approach which would be taken to this problem was also explored. This was stated to be that of a two-stage quantitative modelling procedure. Models of variance between conditions are to be generated (in a data reduction or description stage), and the degree to which the linguistic factors, in which the words differ, are capable of explaining variance in the ERP model is to be assessed (in a data analysis stage). In this chapter I will set out in detail the technique by which this will be accomplished.

Considerable time and energy was required to be invested in the development of a new approach to ERP data reduction and analysis. This was necessitated by difficulties in the application of existing techniques to the present problem. Before proceeding to the technique I will employ, the reasons for deciding not to employ existing techniques are outlined. This is beneficial, since some existing techniques may seem obvious candidates for my purpose, and it is instructive to focus on their specific demerits in this context, in order to avoid such demerits in the chosen approach.

Problems with existing quantitative modelling techniques:

Principal components analysis

Principal components analysis (PCA) (Hotelling 1933; Pearson 1901) takes a matrix of p variables, which are measurements of n items, and finds a new set of orthogonal variables which account for the variance (Kendall 1975). The new set of variables, or principal components (PCs), are ordered so that the first PC explains most variance, the second the next most variance and so on, and it is hoped that the variance in the original matrix will be satisfactorily explained by PCs of number less than p (Manly 1986). The redundancy of the original variables is thus understood to be due to their linear dependence on a smaller set of uncorrelated underlying variables, represented by the extracted PCs (Donchin 1969; Kendall 1975; Wood and McCarthy 1984).

The procedure effected by a PCA is generally explained either by reference to the matrix transformations by which the analysis is numerically computed, or by reference to a (more easily understood) spatial metaphor. The metaphor, which is useful in the explication of most other multivariate techniques (see below), is as follows.

Any data set can be described, without loss of precision, by representing the items as points in a space. The dimensions of the space correspond to the variables on which the items have been measured, and the coordinates of each point correspond to the measurement of the point on each variable. For a typical multivariate data set, the space will have many dimensions (p dimensions), some of which will probably not be orthogonal. The points which represent the items will be distributed within this space in the form of a 'point cloud', which may or may not approximate a multivariate-normal distribution (somewhat like an elliptical galaxy, in three dimensions).

PCA first finds the centroid of the multivariate point cloud, then finds the longest axis along which the points are distributed. This axis corresponds to the dimension in which most variance is expressed. The process then finds the next longest axis in which the points are distributed which is orthogonal to the first. This process is repeated until an arbitrary number of axes are extracted, or the number of extracted components equals the number of the original variables.

It can be seen that this procedure is a geometric re-description of the data into a frame of reference with orthogonal dimensions, and not an inferential statistical process. This is to say, for example, that there is no attempt to dissociate systematic variance from error, or noise, variance. The simplicity, and apparent applicability of the PCA technique to ERP data, made it, for a time, an important means of identifying and measuring ERP components (Donchin, Tueting, Ritter, Kutas and Heffley 1975; Wood and McCarthy 1984). Its extensive use in the field of cognitive ERPs might seem to suggest PCA as an obvious candidate for the present task of generating quantitative models of variance between single word ERPs.

Given the simplicity of the procedure, it is quite surprising that PCA has been demonstrated to misallocate ERP variance to an unacceptable degree in some circumstances (see Chapter 1; Wood and McCarthy 1984). Presently, it is not completely clear why this should be the case (cf., Mocks 1988; Rosler and Manzey 1981; Wood and McCarthy 1984). It is, however, known that misallocation of variance principally occurs between sources of variance which overlap in time (Wood and McCarthy 1984). Since it cannot be guaranteed that the influence of linguistic factors on ERP waveforms will be temporally discrete (see e.g. McClelland 1979), the veracity of any model of variance between conditions produced by PCA could be open to question.

It is instructive to consider another possible problem associated with ERP data structures. It is obvious that ERP data possess inherent sequential dependencies (Wood and McCarthy 1984). This aspect of data structure is itself a problem for PCA, since the geometric transformation effected by PCA is 'blind' to this feature of the data, and hence may mutilate it (Belacicco 1977). What is not obvious (or, perhaps, so obvious that it eludes the eye) is that the form of the sequential dependencies inherent in ERP data will tend to distribute all amplitude measurements along a curvilinear 'backbone'. Whole classes of possible waveform patterns will be absent (e.g. waveforms which are the obverse of the P1-N1-P2 pattern). This means that points in the multivariate point cloud will be clustered together into a curvilinear 'manifold' which curves and twists through the variable space, but which is far from multivariate-normality (Shepard and Carroll 1966).

The absence of whole classes of possible patterns is a diagnostic sign of data structures which are **nonlinear** (Shepard and Carroll 1966). Another, more convincing, sign of nonlinearity is the presence of nonlinear regularities between principal components derived from a data set (Etezadi-Amoli and McDonald 1983). Analysis by PCA of data from the pilot experiment for these studies (see Chapter 3; Young and Routh, unpublished observations) revealed just such nonlinear interactions. It is possible, then, that ERP data structures are sometimes nonlinear. The consequence of this is that ERP data structures are unlikely to be fitted well by a model which assumes a linear dependence of the observed variables on latent ones (Etezadi-Amoli and McDonald 1983; McDonald 1962; 1986; Shepard and Carroll 1966). It therefore seems to me prudent to attempt to avoid possible problems associated with nonlinear ERP data structures in the data analysis strategy.

Principal component analysis can be thought of as Whittle's (1952) equal residual variances common factor model (McDonald 1986). This reflects a close relation between PCA and other methods of linear factor analysis and it was therefore thought unlikely that any other linear common factor model, which would in any case be likely to be untried with ERP data (cf., John, Ruchkin and Vidal 1978), would prove preferable.

Discriminant analysis

This analytic technique addresses the problem of how well it is possible to classify items into two or more groups, given measurements of these items on several variables, and given random samples of items which are known to originate from the groups into which the items must be classified (Kendall 1975; Manly 1986). Hence, it is dependent on prior knowledge of the groups into which items should be categorised. The technique has already been discussed in Chapter 1, in relation to the studies by Chapman and colleagues (Chapman et al 1977; 1978) on the relation between Osgood's semantic differential and ERPs evoked by words sampled from the poles of the differential. The criticism was made there that such a technique only tests the success with which one taxonomic organization (in that case, the organization derived from Osgood's semantic differential) is capable of segregating the data.

The present problem requires that a substantial number of different organizations, derived from the many linguistic factors in which the words could differ, be tested against the ERP data. In this case, then, the only procedure would be to undertake multiple discriminant analyses, one for each linguistic factor, to assess the degree to which each factor is successful in classifying the ERP data.

There are at least two difficulties which militate against this approach. First, there are some technical problems associated with discriminant analyses. These are due to the assumptions that the within group covariance matrix be identical between groups, and that the data within each group follow multivariate-normal distributions (see above; Manly 1986). The technique is unfortunately not robust in the face of violations of these assumptions (Manly 1986). Second, it would be difficult to find a group of words which differ discontinuously in all the possible linguistic factors. The studies of Chapman et al (1977; 1978) exploited the fact that if only one linguistic factor is being compared with ERP data, it is possible to sample words from poles of the factor. It is much more difficult to envisage a stimulus set which consists, for example, only of very long and very short words, very high and very low frequency words, very orthographically regular and very irregular words, and so on, without including words which are at a median position on some other dimension. The factors along which words differ tend to be continuous variables.

For the above two reasons I believe that an approach to the present problem which involved multiple discriminant analyses would be extremely inelegant, and would probably not be feasible.

Multiple regression analysis

Multiple regression is an obvious candidate as a tool for the analyses required, since it is this method which has been employed in psycholinguistic research to address the same problem. There are three difficulties with the application of multiple regression to ERP data which militate against its use.

The first problem is that of 'multicollinearity'. As has been encountered in the behavioural domain (Rubin 1980; Whaley 1978), the linguistic dimensions on which words differ are highly intercorrelated. This feature of linguistic material represents a difficult problem for the application of multiple regression techniques (Cohen and Cohen 1975; Darlington 1968; Edwards 1979; Kerlinger and Pedhazur 1973; Kliegl, Olson and Davidson 1982; Morris 1981; Whaley 1978).

On some accounts (e.g. Cohen and Cohen 1975), the problem of multicollinearity is threefold. First, computer errors are more likely to occur in the computation of the regression analysis if any independent variables are highly intercorrelated (indeed, the computation may be impossible: Edwards 1979, p 65). Second, it may be difficult to provide satisfactorily stable estimates of the regression coefficients for substantially intercorrelated variables (see also Whaley 1978). Third, the interpretation of the regression coefficients will necessarily be misleading (Cohen and Cohen 1975; Whaley 1978) in the case that two or more predictor variables share variance.

There are some methods for countering the difficulty represented by multicollinearity, the most common of which is to submit the predictor variables to a principal components analysis prior to attempting a regression (Darlington 1968; Edwards 1979; Kerlinger and Perdhazur 1973; Rubin 1980; Whaley 1978). There is little doubt that this procedure, which guarantees that the variables used in the regression are uncorrelated, yields a solution with statistical propriety. But there is no guarantee that the computed principal components will be straightforwardly related to the sets of variables in which the experimenter is interested. For instance, a particular principal component may relate closely to, say, word frequency variables, may also relate to letter frequency variables, and may be somewhat related to orthographic variables. It would be difficult to interpret the finding that such a PC explained variance in the dependent variable in terms which are generalisable beyond the particular data (i.e. in terms of the original variables). Problems with interpretation of regressions based on principal component variables are only mitigated if the extracted factors conveniently fit a taxonomy which the experimenter can apply (e.g. factor 1 relates, say, to all the word frequency variables, factor 2 to interletter probability structure and so on). This has occasionally been the case (e.g. Whaley 1978).

The second difficulty is that all the multiple regression models, of which I am aware, take a univariate dependent variable. Variance between ERP conditions, however, cannot be guaranteed to be expressible in only one dimension; as for example, in the case that a simple DC shift is present between traces at the majority of electrodes. If variance between ERPs takes the form of complex differences in the statio-temporal distribution of the scalp potential field, as has been suggested is the case for verb/noun interpretations of homophones (Brown, Lehmann and Marsh 1980), representation of the structure of differences between conditions may require more than one dimension. In fact I had some expectations, grounded on empirical (see Chapter 3) and theoretical (see Chapter 8) results, that at some latencies a unidimensional model would not adequately capture the complexity of the differences between ERPs recorded to individual words. For this reason, the constraint imposed by multiple regression, that the dependent variable be unidimensional, was considered unacceptable.

A third problem with the application of multiple regression techniques to ERP data, in the present context, is that the 'power' of regression analysis is related to the number of measurements in the dependent and independent variables. For example, the regression studies of Whaley (1978) and Rubin (1980) measured RT to 100 and 125 words, respectively, each of which had been characterised on a number of linguistic dimensions. To employ regression analysis here, it would be necessary to measure brain potentials to a large number of words. Because the measurements in question involve the evocation of an ERP to each word, such a procedure would entail experiments of impracticable length. If I assume that an ERP condition should be evoked by no less than 40 trials (to allow for loss of trials due to eye-movement artifacts or behavioural errors) and further assume that a regression analysis requires 50 words (half that deemed desirable by others: e.g. Rubin 1980; Whaley 1978), it can be seen that the number of trials in the experiment (not counting any other conditions) would be 40 times 50 trials, or 2000 trials. Experiments in this laboratory, conditioned by standards of appropriate treatment of volunteer subjects, rarely exceed a quarter of that length. The time constraints imposed by the necessity for averaging, in order to derive an ERP, make an application of regression analysis to the present problem very difficult to envisage.

These three problems, the problem of intercorrelated variables, of the constraint that the dependent variable be unidimensional, and the problem of impracticably long experiments, were considered effectively to rule out the use of regression analysis to address the relation of ERPs to linguistic factors.

A new quantitative modelling approach

The field of event-related potentials has seen the introduction, rise and subsequent fall of a substantial number of statistical procedures. It is with a note of caution, therefore, that I introduce yet another one.

The problems associated with existing techniques which dissuaded me from their use were overviewed above. Part of the problem with the variety of procedures employed with varying success in this field, may be that ERP data-structures are rather idiosyncratic. ERP data exhibit time dependency (Wood and McCarthy 1984), they may exhibit nonlinearity (see above), and there are often overlapping sources of variance manifest in particular voltage deflections (Curry and McCallum 1982). Further there are problems for some of the assumptions of classical parametric techniques such as ANOVA (Keselman and Rogan 1980). Indeed, there are problems of experimental design and analysis associated with the ERP methodology itself: sharp constraints of time are imposed by the necessity for averaging to obtain the waveforms.

In order to be a candidate for the present task it is desirable, therefore, that a statistical procedure should meet the following criteria:

(1) It should require few conditions to identify the best predictors of variance (In order to allow experiments of practicable length).

(2) It should allow mathematical comparison between numerical models of different dimensionality (in order to allow linguistic factors which may have different dimensionalities to be simultaneously compared to the ERP variance model, see Chapters 3 to 5).

(3) It should not require the linguistic predictor variables (factors) to be orthogonal (because for naturally occurring language materials the factors are not orthogonal, and there are interpretative problems associated with analyses which assume that they are).

(4) It should avoid possible problems of nonlinearity in ERP data structures.

(5) It should make as realistic assumptions as possible about ERP measurement, and should not misallocate variance between aspects of the mathematical representation of the ERP data (as does principal components analysis, Wood and McCarthy 1984).

(6) It should not make assumptions about the characteristics of ERP data which are unlikely to be met (e.g. multivariate normality, as is assumed by parametric techniques).

The analytic procedure set out below, I believe, meets most of these criteria. I have tailored it to the characteristics of ERP data, as I see them, and the particular statistical problem generated by the decision to address the central question identified in Chapter 1. My presentation of it will take the form, first, of a piecemeal description of the component techniques, and then an overview of the whole statistical process.

Profile comparison

There are two equivalent ways in which to represent the interrelations between items of which measurements have been made. The first is to represent the items as **multivariate data** in the form of an n x p matrix, where n denotes the number of items and p the number of variables, (or dimensions) on which the items have been measured. The second is to represent the items as **proximity data** in the form of a matrix of differences between the items, that is, in an n x n matrix. These two representations are interchangeable. In the case of the conversion of multivariate data to proximity data, this is achieved by computing all-pairwise 'distance' statistics between conditions. In the case of the conversion of proximity data to a corresponding multivariate representation, the interchange is brought about by submitting the proximity data to multidimensional scaling (Kruskal 1964; Shepard 1962; 1972; 1980).

ERP data are, of course, multivariate data: each condition is represented by amplitudes measured at each electrode and time-point. Almost all the approaches to ERP data analysis which have been attempted hitherto have remained within this data format. The translation, however, of (any kind of) multivariate data into corresponding proximity data opens the door to a large variety of statistical procedures. These include the varieties of cluster analysis

(Aldenderfer and Blashfield 1984; Cormack 1971; Dunn and Everitt 1982) and the family of analytic techniques known as multidimensional scaling (MDS) and multidimensional unfolding (Shepard 1980; Shepard, Nerlove and Romney 1972).

In particular, this interchangeability of multivariate and proximity representations of the items in a data set can be exploited to provide a means of data reduction and description. The first statistical problem that any approach to ERP data analysis must encounter is that of finding a parsimonious, but accurate, description of the data. This problem can be construed as that of dimensional reduction (Donchin 1969). By this is meant that the large number of original variables on which the conditions are measured must be reduced to some smaller set of variables. This small set of variables should have the threefold virtue of being understandable, mathematically tractable, and of faithfully representing the data structure.

This process of dimensional reduction can be achieved by converting the multivariate ERP data to proximity data, and then employing MDS to produce an optimally fitting multivariate description of the data in a small number of dimensions.

As an aid to comprehension of this process, again imagine that each item is represented as a point in a high dimensional space. The dimensions of the space are the large number of original variables on which measurements of the items have been made. The coordinates of each point in this space are given by the value of the measurement (in this case, the amplitude) of each item on each of the dimensions. Now, despite the complexity of the space, in terms of its large number of dimensions (indeed, independent of the number of dimensions), the points will be more or less close to one another. Points close together have similar measurements on the dimensions, and points far apart have dissimilar measurements. By measuring the distances between points in this high-dimensional space, one can derive a representation of their relationship one to another. These distances capture the relations between conditions in the same way that an intercity distance table captures the relations between cities (which are distributed in space). This is analogous to the derivation of proximity data from multivariate data.

The next step involves adjusting the arrangement of points in a space of specified low dimensionality until the distances between them mirror as closely as possible the distances between the points in the high-dimensional space (Kendall 1975; Manly 1986; Shepard 1962; 1980). This is achieved by minimizing a cost function which is related to 'badness of fit' between the set of distances relating to the high-dimensional space, and the set of distances relating to the solution space (Kruskal 1964). The 'cost' is similar to residual variance-explained in a regression analysis (Manly 1986). The configuration of the points as measured on the small number of dimensions selected for the solution is then output. This process of fitting a low-dimensional configuration, by minimizing the mismatch between the distances between points in the high and low dimensional spaces, is that accomplished by multidimensional scaling.

The first step in the process of generating the required variance models, then, is the reduction of the multivariate ERP data to proximity data. The distance statistics between conditions are derived by 'Euclidean distance' profile comparison:

$$d_{(x,y)} = \left(\sum_{i}^{t} (X_i - Y_i)^2 \right)^{1/2}$$

Where X and Y are two conditions between which the euclidean distance $d_{(x,y)}$ is computed, and i is an index of the time-points of a latency window of length 1 to t.

This procedure is equivalent to measuring the straight-line (geodesic) distance between points in the multivariate point cloud. In terms more related to the ERP data, for each electrode, the profile comparison simply sums the squared differences between homologous points in the chosen latency window, and gives the square root of the summed differences as the distance statistic.

This widely used procedure (being the default in the appropriate routines of the SPSS-X statistical package) has several advantages. It cannot misallocate variance, as does principal components analysis, as there are no latent variables between which variance may be misattributed. It sidesteps possible problems associated with nonlinear data-structures by being dependent only on the distances between points in the original variable space (Shepard and Carroll 1966), rather than being directly dependent on the characteristics of the space itself. For example, even in the case that the multivariate point cloud is concentrated in a curvilinear form (which will be misdescribed by PCA, see above; McDonald 1987; Shepard and Carroll 1966) the distances between points will be approximated by a euclidean distance (see Ashby and Perrin 1989; Shepard 1972). The profile comparison procedure is also very simple: regions of the waveforms which differ contribute to the distance statistic and those that do not, correspondingly do not.

The employment of difference waveforms is a widespread tool in the field of psychophysiology (e.g. Bentin, McCarthy and Wood 1985; Lovrich, Novick and Vaughan 1988; Naatanen, Gaillard and Mantysalo 1978; Ritter, Simson, Vaughan and Macht 1982; Kutas and Hillyard 1980a; 1980b; Rugg 1987). The euclidean distance metric (being the root-sum-of-squares) varies as the root-mean-square of the difference waveform between two conditions. Very similar kinds of area measure have been employed to examine differences between ERP waveforms (e.g. Kutas and Hillyard 1980a), though probably without the knowledge that such measures are, in fact, closely related to the widely used euclidean distance. The euclidean distance metric seems, therefore, to be a reasonable approach to the quantification of differences between conditions, since it involves no more complex or novel assumptions concerning differences between waveforms than are routinely employed in this field. This is particularly defensible when it is recommended that "it is probably reasonable to say that any measurement or quantification technique which produces reliable, interpretable data is acceptable." (Rugg et al 1986 p275).

The euclidean distance metric is to be contrasted with metrics based on correlations between conditions. Correlation measures will tend to underestimate differences in the case that a sustained (low frequency or D.C.) difference is present between traces which otherwise possess a similar morphology (Donchin 1969). The euclidean distance metric is, obviously and by contrast, sensitive to this type of ERP difference. It has been asserted, however, that "for normalised ERPs, the euclidean distance is equivalent to the Pearson correlation coefficient and constitutes a measure of waveshape similarity" (John, Ruchkin and Vidal 1978 p159). This is not quite correct. The euclidean distance equals the correlation measure only when the values of the waveforms *in the latency window* which is submitted to an analysis are normalised. This practice was never carried out in the present work, and all waveforms were 'normalised' to the average of a 100 msec prestimulus baseline (see Chapter 3). Hence, the euclidean distance procedure used here is inequivalent to a correlation measure.

Multidimensional scaling

The above procedure, then, derives proximity matrices between each condition. In order precisely to convey the next step in the data reduction process, I now need to describe the format of the ERP data in more detail.

For each subject, amplitude measurements will be made at a number of electrodes, at each time-point of the post-stimulus epoch, and will correspond to each of the set of presented words. Thus, each subject's data will consist of a cubic matrix whose dimensions are conditions (words) by electrodes by time-points.

The simplest approach to finding a representation of the interrelations between conditions would be first to collapse the cubic matrix into a two dimensional n x p matrix. This could be done by 'stacking' the electrode variables into a long series of electrode-and-time-point variables. For example, condition 1 would be represented in the new matrix by a single set of variables corresponding to the time-points associated with electrode 1, then the time-points associated with electrode 2 and so on. This two-dimensional matrix could then be submitted to profile comparison and a proximity matrix generated on the basis of the combined electrode and time-point variables.

This approach would entail a brisk data reduction from the full data set for each subject (consisting of more than 30,000 numbers in the experiments which follow) to a single proximity matrix of only n(n-1)/2 numbers. The proximity matrix for an experiment based on 10 conditions would thus contain 45 numbers.

The proximity matrix could then be submitted to a simple MDS program, such as Kruskal's MDSCAL (1964) to derive the quantitative model. This would have the advantage that the 'badness-of-fit' statistic (called Stress) produced by the program, has been the subject of a number of Monte Carlo studies (e.g. Stenson and Knoll 1969). These were undertaken to determine the stress values which would be associated with purely random data, and can hence be used as a guide to the likelihood that the proximity matrix contains systematic information (Stenson and Knoll 1969).

Unfortunately, this approach is unworkable for the following reasons. The proximity matrix for 10 conditions contains 45 parameters. To make as few assumptions about the characteristics of the original variables as possible it is assumed that the proximity statistics are at no higher level of measurement than ordinal (rank) data (Coombs 1964; Shepard 1972). For a unidimensional MDS solution, 10 parameters would be extracted from this input data, corresponding to the coordinates of the ten points on this one axis. For a two dimensional solution, 20 parameters would be extracted, and for a three dimensional solution 30 parameters and so on. It can be seen that with increasing dimensionality the extracted, metric, parameters approach (and will eventually exceed) the number of input parameters. As the output parameters approach the number of input parameters, precise estimation of the former becomes increasingly difficult.

This problem is sometimes called the problem of 'metric indeterminacy' and is a consequence of the small number of conditions relative to their possible dimensionality. It is difficult to envisage a means to deal with the problem within the above approach, as the number of conditions are constrained to be small by the necessity for averaging, and solutions of dimensionality more than one were specifically sought (see the above section on multiple regression).

The process of data reduction used here is slightly more complex than the simplest solution discussed above. Rather than submitting a two dimensional data matrix to profile comparison and subsequent MDS, the method retains the cubic data matrix (of conditions by electrodes by time-points) to describe each subject's data. This data matrix is submitted to the profile comparison procedure in such a way that it generates a proximity matrix between conditions *for each electrode*. Thus, instead of a single proximity matrix, there are a number of them. For the first several experiments (see Chapter 3), for example, there are 13 electrodes, each of which will give rise to a proximity matrix of 45 numbers, which express the similarities and differences between the conditions as 'seen' by each electrode. The first stage of data reduction is, hence, from the order of 30,000 numbers to about 600 numbers.

A metric coordinate structure representing variance between conditions is then generated by three-way nonmetric multidimensional scaling (INDSCAL: Carroll and Chang 1974). This model represents, for a subject, the interrelations between ERPs evoked by exposure to the different words. The analysis also provides weights which signal the salience of each output dimension for each electrode channel (i.e. how well an electrode 'told the difference' between the extrema of each dimension). These electrode weights are themselves interpretable, in the same way that weights derived from different subjects' judgements have been interpretable (Wish, Deutsch and Biener 1972). They might indicate, for example, which scalp electrodes contribute most to differences between the words, on a dimension which can be identified as being related to one of the linguistic factors.

The above procedure derives a variance model which represents the interrelations between ERPs evoked by the words *for each subject*. Structures for each subject are entered as data sources into a further profile comparison and MDS routine, to derive a structure which represents the interrelation between evoked potentials for all subjects: that is, the required quantitative model of variance between conditions.

This process involves, again, the derivation of a proximity matrix, but this time from the multivariate data in each subject's variance model. The proximity data for each subject are then submitted to INDSCAL in the same way that proximity data for each electrode were submitted in the within subject analysis. The output configuration thus expresses the structure of differences between the ERPs evoked by each of the words in the presented set. It does so on the basis of all the amplitude measurements at each electrode of each subject, and at every time-point of the latency windows which are the subjects of analysis.

Because of the presence of noise in the waveforms, which is not segregated from signal by this technique, each point in the metric structures should be understood as being only *representative* of the 'real' (noise-free) point. This 'real' point will be somewhere within an 'error-sphere' (if the solution has three dimensions), located around the computed point, whose diameter may be quite large. This further round of profile comparison and MDS may improve the signal to noise ratio of the resultant, global, variance structure over that obtaining in each subject's structure. There is reason to hope, then, that the error-spheres present in the final solution may be smaller than those in the intermediate analyses, and that the final configuration of computed points is thus more representative of the ’real’ points. The rationale for this is that the final model fits those aspects of the within subject models which agree (signal), and loses those aspects which do not (noise).

Note that collapse across subjects takes place at the level of their variance models and not at the level of their waveforms. The actual distribution of potential between conditions which gives rise to each subject’s model need not be assumed to be the same. It is known that there are gross morphological differences between brains, for example in their expression of the primary visual area (Braak 1978). The possibility cannot be excluded that these morphological differences may be reflected in variant patterns of the distribution of potential over the head. For example, a brain in which foveal visual representation was accomplished by cell populations on the bank of the calcarine fissure could give rise to a different potential distribution, on checkerboard reversal, than one in which the homologous population was situated on a nearby gyrus. Both potential distributions would nonetheless reflect checkerboard reversal. Different potential distributions might also, by analogy, reflect the influence of the same linguistic factor. This possibility is allowed by the technique used here, in contrast to ANOVA models on conventional amplitude measures in which such variance would be fitted as noise.

PROCRUSTES rotation

The global ERP variance model, derived as above, is then compared with models designed to capture the ways in which the words differ according to each of the linguistic factors. These linguistic models are derived from subjective judgement experiments based on the presented words, and from normative data, such as word and letter frequency tables. This comparison, which is the core of the data analysis process, is achieved by means of PROCRUSTES rotation.

PROCRUSTES is a least squares method of assessing the configurational similarity of two n x p matrices (n = items, p = dimensions), where the number of dimensions of the two may be different. Since the ERP model and the linguistic models are determined only "to within a similarity transformation", the representations must be transformed into mutual congruence before their agreement may be assessed (Shepard 1972). This transformation is accomplished by finding the reflection, rotation or homogeneous scaling (stretching, centred on the centroid, which is equal for all dimensions) which minimises the difference between the two structures. The routine then provides variance-explained and residual variance statistics, which describe the goodness of fit between the ERP model and each of the linguistic models against which it is rotated. Thus the outcome of this part of the data analysis procedure is a table of variance-explained statistics, which details how well variance expressed in an ERP model is accounted for by each of the linguistic factors.

The configurations produced by MDS are scaled in such a way that the recovered dimensions possess about the same variance. Some dimensions, however, are more 'important' in that they account for more of the input variance than other dimensions. This difference of 'importance' is represented by the 'overall importance of each dimension' statistics which are produced by the MDS program. These statistics sum to the mean variance explained value for a solution. The scaling effected by PROCRUSTES, when two configurations are compared, is homogeneous (see above). In order to preserve the relative importance of each dimension and to avoid giving 'unimportant' dimensions more salience than they actually possessed, the dimensions of each configuration were multiplied by the 'importance' statistic associated with it. In this way the variance of each dimension of the configurations were scaled to accord with their relative contribution to explaining variance in the input data.

One advantage of using the PROCRUSTES procedure is that it is able to compare models of different dimensionality. As I have already discussed, and as will be apparent in Chapters 3 to 5, this is an essential feature of the data analysis process.

Another approach to the comparison of models of different dimensionalities is to assess the similarity between their corresponding proximity matrices (Manly 1986). Entries in a matrix of proximities, however, are not statistically independent measurements because they relate to the same objects. Correlation alone is therefore not satisfactory (see Chapter 3). An approach due to Mantel (1967; Manly 1986; Mielke 1978), while similar to correlation, can assess the rarity of a relation between two distance matrices. This is achieved by first computing the cross-products of the two proximity matrices (Mantel 1967), or their correlation (Manly 1986). The objects (i.e. the 'labels') to which entries in one of the matrices relate, are then permuted, and the statistic calculated on each permutation for a representatively large number of permutations. If the observed statistic is unusually large, then larger statistics are unlikely to be produced by the randomized tests. This likelihood can be expressed as a probability that the observed statistic was due to chance.

The disadvantage of Mantel's test, as compared to the use of PROCRUSTES, is that the Mantel statistic itself (or correlation, as above) is not particularly interpretable. PROCRUSTES produces a variance-explained statistic for each of the linguistic models which is interpretable. The Mantel procedure is, however, similar to the final procedure in the data analysis approach taken here (see below).

The PROCRUSTES routine used was that instantiated in the GENSTAT statistical package (Alvey 1980). There have been several related methods developed for the purpose of transforming two structures into mutual congruence and assessing their agreement (Carroll 1968; Cliff 1966; Lingoes 1967; Schonemann 1966; Schonemann and Carroll 1970). The specific method chosen by the GENSTAT programmers is that of Schonemann and Carroll (1970), as further explored by Gower (1971).

Randomization testing

Mantel's test, noted above, is one of a general class of statistical procedures which are known as randomization tests (Edgington 1980). Almost any statistical test may be subjected to a randomization test in order to have the rarity of an observed statistic assessed (indeed, this is the origin of the values in statistical tables for idealised data (Fisher 1936)). This may be done in either of two ways. The first is by systematically permuting the data through all n! possible orderings, so that a distribution of the test statistic for the data in question may be computed, and the position of the observed statistic compared with the distribution. The second is by counting the number of times that the test statistic is exceeded by random permutations of the data (Edgington 1980).

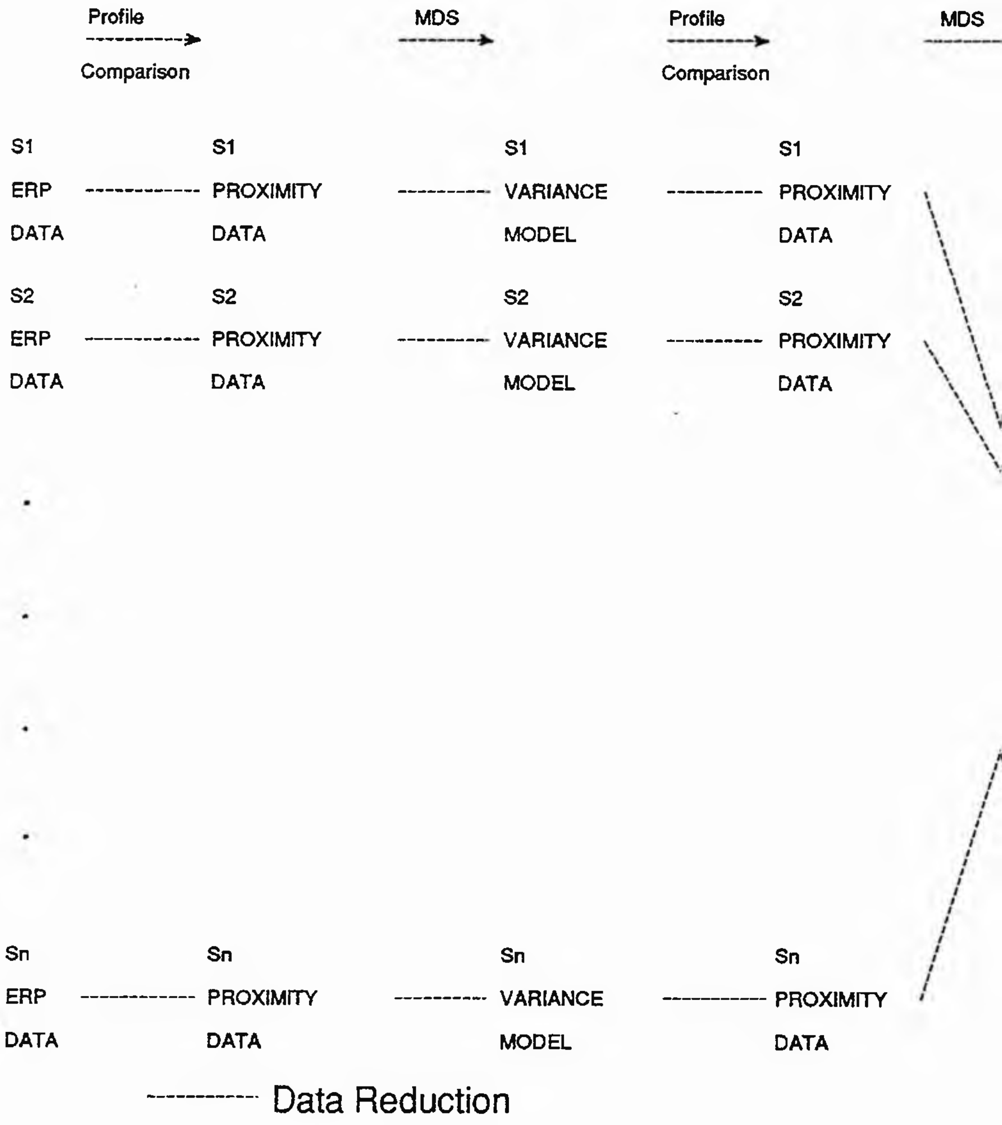
This second approach is less demanding of computer resources. A representative sample of the n! possible permutations can be used (a so-called approximate randomization test (Edgington 1980)), rather than a systematic randomization test involving every possible permutation. If an approximate randomization test is undertaken with sufficiently many permutations, the computed probability will closely reflect the probability derivable by systematic randomization (Edgington 1980).

So, as the final step in the statistical process, the rarity of each variance explained statistic is assessed by using an approximate randomization test based on Fisher's method of randomization (Edgington 1980). This involves permuting the labels of each variance model 400 times (a random sample of the n! possible permutations), while computing a variance-explained statistic by PROCRUSTES rotation each time. This computationally intensive procedure yields an approximate rarity value, expressed as a probability that the observed value could have been obtained by chance. This probability indicates how seriously the corresponding variance-explained statistic should be taken; a low probability, high variance-explained relation of a linguistic model is less likely to have come about by chance than a high probability, low variance-explained value. It is important to note, though, that this 'significance' is not related to estimates of population parameters as in ANOVA.

The whole process: Profile comparison, MDS, PROCRUSTES, and randomization testing

The data analysis scheme is depicted in Figure 2.1.

The data reduction and description procedure employed here thus consists of several stages. First, the large amounts of ERP data ($\sim 10^5$ numbers)are reduced by the derivation of proximity matrices for each electrode($\sim 10^3$ numbers). Quantitative models are then generated for each subject in a further data reduction step which collapses across electrodes ($\sim 10^2$ numbers). The subjects' models are then converted to proximities once more ($\sim 10^2$ numbers). The resulting matrices are fitted by MDS to produce a global variance model which best fits the data, in a final data reduction step which collapses across subjects ($\sim 10^1$ numbers).



The data reduction and analysis procedures

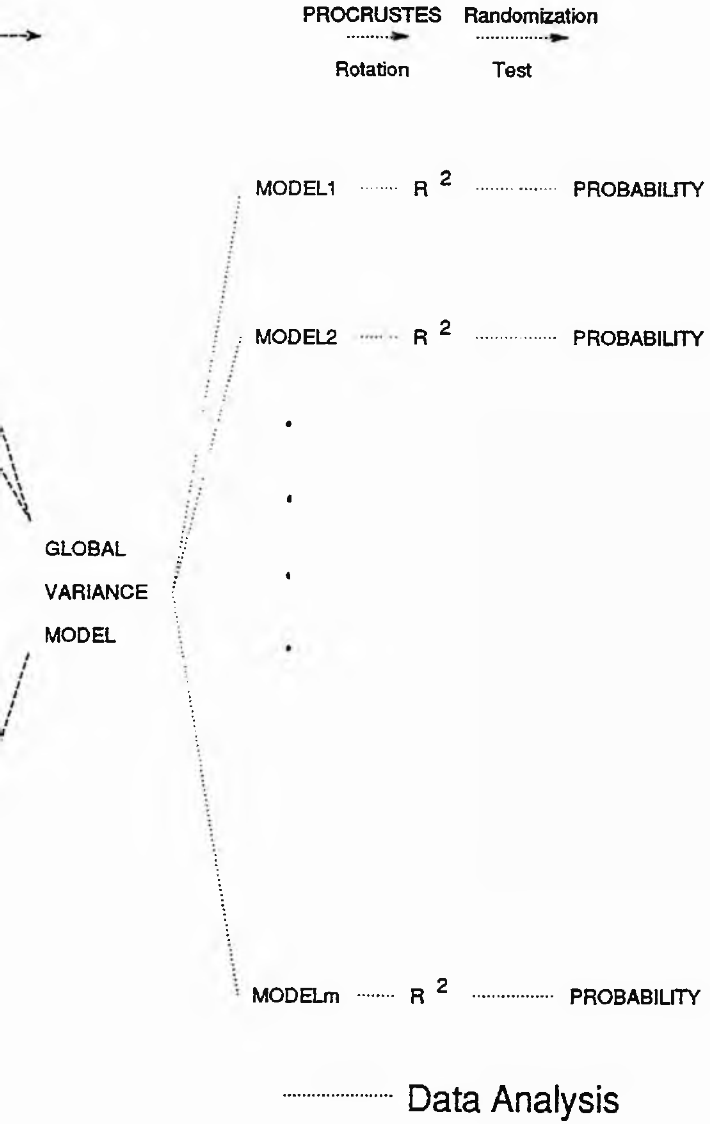

PROCRUSTES
Rotation
Randomization
Test
MODEL1
R 2
PROBABILITY
MODEL2
R 2
PROBABILITY
GLOBAL
VARIANCE
MODEL
MODELm
R 2
PROBABILITY
Data Analysis

FIGURE 2.1

The processes of PROCRUSTES rotation of each linguistic model with the ERP model, and the subsequent randomization test, together comprise the data analysis part of my statistical approach. The outcome of the application of these techniques is not dissimilar to the outcome of the regression analyses undertaken by psycholinguists, when the same problem was addressed in the behavioural domain (e.g. Rubin 1980; Whaley 1978). The results take the form of a table of variance-explained values relating to each linguistic factor, and the corresponding probabilities of their having been obtained by chance.

Taking the data reduction and data analysis procedures together, the approach could be described as a kind of **configurational analysis**. The central aim is to derive a configuration which represents **the structure of differences** between ERPs evoked by individual words, and then to test the similarity of this structure to configurations which represent the ways in which the words would differ if the linguistic factor which they embody were responsible for the ERP differences.

Notwithstanding my attempts to derive a method which is well suited to ERP data, which is statistically defensible, and which is capable of answering the question which I have chosen to address, there remain a number of practical problems which should be borne in mind.

The first practical problem is that the basis of the technique is the identification of linguistic factors by means of a measure of variance-explained. Inevitably some unknown proportion of the variance in the ERP model must be due to noise. I noted above that the collapse across subjects is likely to improve the signal to noise ratio in the global model over that in the within subject models. But it is still likely that the global ERP models will be perturbed by noise. The linguistic models are likely to be able to explain less variance in these noisy, empirical models than in some hypothetically noise-free ERP model.

The second set of practical difficulties is derived from the necessity for averaging in order to derive an ERP. It cannot be guaranteed that the various linguistic factors will affect ERP variance homogeneously over the time course of an experiment (see also Rugg, in press). It might be the case, for example, that word frequency effects are present in the earlier trials, but tend to diminish during the multiple repetitions of the stimuli as the experiment proceeds. In consequence, frequency effects might be under-represented in the waveforms. Indeed, in advance of the facts, this might be expected on the basis of the attenuation of word frequency effects by repetition in behavioural studies (e.g. Forster and Davies 1984). Similarly, the possibility exists that there may be 'semantic saturation' effects due to the multiple repetition of the stimuli (Boddy 1981). Thus, the experimental situation which is enforced by ERP methodology, is one in which the influence of linguistic factors on the ERP would be expected to be difficult to detect.

A third practical problem relates to the interpretation of p-values associated with particular linguistic models. An important distinction in the design and analysis of psychophysiological experiments has been drawn between research which can be characterised as confirmatory, and research better thought of as exploratory (Muller, Otto and Benignus 1983). As the title of this thesis emphasises, the basic motivation of the present research is exploratory. In the case of exploratory research, individual p-values should be interpreted with caution (Muller, Otto and Benignus 1983).

The reason for this is the Bonferroni inequality (Abt 1981; Kirk 1969), which relates the probability of type I error (deciding that there is a reliable effect in its absence) to the nominal significance level multiplied by the number of independent tests. As the number of possible factors against which the ERP model will be tested increases, so does the probability that a p-value will descend below the nominal (0.05, say) significance level by chance. Adjusting the significance level downwards reduces the probability of type I error at the cost of increasing the probability of type II error (deciding that there is no reliable effect in its presence) (Muller, Otto and Benignus 1983). Two responses are called for. First, as above, the p-values should be taken only as a guide to the degree of seriousness with which an associated variance-explained statistic should be viewed. Second, a greater relative emphasis should be placed on replicability, on split sample techniques, and on the search for regularities in the patterns of variance-explained statistics across different subject groups (Muller, Otto and Benignus 1983).

A fourth practical problem is that the statistical procedures which comprise the data reduction and analysis strategy are very costly indeed in terms of computer power. This is particularly true of the randomization tests because they involve many PROCRUSTES rotations, each of which is computationally intensive in itself. The consequence of this problem, at the present level of computing technology, is that the approach is very much more time consuming to implement than a conventional approach employing a classical parametric omnibus test, such as ANOVA.

A fifth concern, which may or may not be a difficulty, is that problems with the application of principal components analysis to ERP data were uncovered only by extensive simulation testing with data whose parameters were known (e.g. Wood and McCarthy 1984). Might it not also be the case that the novel analytic technique developed and applied here suffers from some similar subtle demerit, which could only be rooted out by similar extensive testing?

The analogy between the two techniques and their applicability to ERP data, however, should be treated with caution. There are four points to raise in this context. First, the difficulties associated with the application of PCA to certain kinds of data structure, of which ERP data may be an example (see above in relation to PCA), have been known for more than 20 years (e.g. McDonald 1962; Shepard and Carroll 1966): that detailed testing was required to point up the demerits of PCA may, in part, reflect the fact that this literature was unknown to those concerned with methodological aspects of psychophysiology. Second, the specific problem associated with PCA, of misallocation of variance between components (Wood and McCarthy 1984), simply does not apply to the present analyses. There are no latent variables here, between which variance could be misattributed (see above, in relation to profile comparison). Indeed, the analyses are not concerned with variance between components, but with quantifying and representing variance between conditions. Third, extensive testing of the MDS routines used for data reduction has been undertaken (Shepard 1962; 1972; 1980; Stenson and Knoll 1969). Shepard (1980 p 392) notes that "a test configuration of as many as 15 random points could be essentially reconstructed ... and with as many as 45 random points, I later found that product-moment correlations between true and recovered distances averaged over 0.9999997". Hence, it is very likely that the variance models reliably reflect the proximity matrices which they receive as input. Concerns about misrepresentation of the data by the data reduction process therefore converge on the issue of the adequacy of euclidean distance profile comparison to the task of quantifying differences between conditions. This I have already discussed above. Fourth, simulation testing is one of two responses which could be made to the uncertainties associated with a novel statistical approach. The other is to test the results of the modelling process by means of conventionally designed and analysed experiments. As discussed in Chapter 1, this latter course is the one I have chosen. The grounds for this choice were that the identification of the linguistic factors in ERPs evoked by language materials was my prime concern, and the necessarily complex method employed to accomplish this was secondary: empirical testing allowed the work to address empirical issues, rather than to be an arid treatise in multivariate statistics.

The new approach to ERP data analysis taken here, both in terms of the actual processes, and in terms of the general motivation, is rather different to that undertaken elsewhere in cognitive psychophysiology. In pursuing this unconventional course, the remarks of Tukey (1978 p 165) were heartening: "The man or woman who seeks to make important contributions, and who sees .. where a problem may be attacked, will be ineffective if the way in which the problem is attacked is too much controlled by a combination of: 1] the approaches customary in the field, and 2] the most common techniques of data analysis. There is no substitute for thinking through (a) what one really wants to do, (b) how the problem should be approached to do this, and (c) how the resulting data should be approached in its analysis. New questions often require new techniques for their successful answering". I cannot aspire to 'make important contributions', of course, but the approach I have taken is intended to exemplarise Tukey's suggestion that well motivated innovation is to be preferred to the inappropriate application of conventional knowledge.

Chapter Three

An electrophysiological and behavioural investigation of the colour names

General Introduction

This chapter presents the empirical results of the application of the quantitative modelling technique described in Chapter 2. The method is applied to individual words chosen from the category of colour names. This category was selected because aspects of the psychological properties of the colour words have been extensively investigated.

Maerz and Paul (1930), in their 'Dictionary of Color', detail more than 3,000 English colour terms. The colour names in this very large lexical category vary widely in their frequency, familiarity and utility in every-day discourse. Indeed, there is evidence from confusions of colour naming (Halsey and Chapanis 1954), that, underlying this plethora of possible terms, the domain which they describe may consist of only 11 to 15 reliably identifiable colours (Miller and Johnson-Laird 1976).

In an attempt to place an organization on the colour domain, thirteen colour names; RED, ORANGE, YELLOW, GREEN, BLUE, VIOLET, PURPLE, PINK, BROWN, OLIVE, BLACK, GREY and WHITE, were chosen by the Inter-Society Color Council of the (U.S.) National Bureau of Standards (ISCC-NBS) to comprise the basic terminology. Secondary or intermediate colours are defined in relation to these terms, which denote the 'hue' of the colour. Particular colour experiences may be described by the addition of '-ish' (e.g. greenish blue) and by degrees of saturation ('washed-out' to vivid colour) and of lightness (dark to light colour). Thus, the ISCC-NBS system of colour naming can be conceptualised as a three dimensional 'colour solid', with a shape similar to that of two cones joined at their bases. The dimension from one cone tip to the other corresponds to lightness (Black at one tip and white at the other), the dimension from the centre of the structure to the circumference of the 'joined bases' corresponds to saturation (grey at the centre and vivid colour at the circumference), and the gradation around the circumference corresponds to the hue (the so-called colour circle).

The landmark colours chosen by the ISCC-NBS, accord well with the colour descriptors in common use. Battig and Montague (1969), for example, found that when a large number of college students were required to write down (in 30 seconds) as many colour names as they could recall, twelve colour names were most frequently produced. In decreasing order of production frequency, the twelve colour words were; BLUE, RED, GREEN, YELLOW, ORANGE, BLACK, PURPLE, WHITE, PINK, BROWN, VIOLET AND GREY . This order has all the ISCC-NBS basic terms except OLIVE (Miller and Johnson-Laird 1976).

Notwithstanding the attempt by the ISCC-NBS to place a generally agreeable organization on the colour domain, there is some disagreement concerning the basic, landmark, colour terms. Miller and Johnson-Laird (1976), for example, dispute that OLIVE and VIOLET are landmark terms on the ground that neither will (felicitously) take the suffix '-ish'. This leaves the eleven items selected by Berlin and Kay (1969) as the basic colour terms: RED, ORANGE, YELLOW, GREEN, BLUE, PURPLE, BROWN, PINK, GREY, BLACK AND

WHITE. Miller and Johnson-Laird (1976) further reduce the list to a 'minimal constrastive set' by exclusion of BROWN (any warm colour of low saturation), PINK (light red), ORANGE (more commonly a fruit than a colour) and PURPLE (really the purpura dye). The resulting six colour names: RED, YELLOW, GREEN, BLUE, BLACK and WHITE, probably capture economically the psychologically most salient aspects of colour (Miller and Johnson-Laird 1976).

The application of multidimensional scaling (MDS) techniques to judgements of the perceptual similarity of presented colour patches confirms that the basic organization of the colour domain is well approximated by the three dimensional colour solid described above (Indow and Kanazawa 1960; Indow and Uchizono 1960; Miller and Johnson-Laird 1976; Shepard 1962; 1972; 1980; Shepard and Carroll 1966). The application of MDS to judgements of the *semantic* similarity of the colour words (with nothing more colourful in the presentation than black ink on white paper, or white letters on a dark-grey computer screen) gives a similar spatial arrangement to that recovered for perceptual judgements (Fillenbaum and Rapoport 1971; Rapoport and Fillenbaum 1972).

There are, however, some differences between the perceptual and semantic structures. The lightness and hue dimensions tend to be orthogonal for perceptual judgements. But the structure recovered from judgements of semantic similarity of colour words has oblique axes between these two dimensions. This is to say that the polarity between BLACK and WHITE is not completely dissociated from the polarity between hues, for example that between BLUE and YELLOW (see below; Young and Routh, unpublished observations). A possible explanation for this is that judgements of semantic similarity are made on the basis of memory for the most typical colour represented by each of the names (see also Miller and Johnson-Laird 1976). Because the most typical YELLOW is a light colour, and the typical BLUE is a dark colour, these colours are judged more similar to WHITE and BLACK, respectively, than are judgements of the perceptual similarity of (atypical) blue and yellow colour patches. The structure derived from semantic judgements by MDS thus shows the colour circle canted away from a plane perpendicular to the black-white dimension.

Another feature of the configurations derived from semantic similarity judgements of the colour names is that those of colour-blind subjects may not differ from those of normal subjects (Young and Routh, unpublished observations). This finding is uncomfortable for the supposition that the perceptual and semantic configurations are simply related. It suggests that the representations, on the basis of which semantic judgements are made, need not be very closely related to perceptual representations: colour-blind subjects have no perceptual experience of some of the distinctions between colour names which they are asked to evaluate in these experiments. They do, however, experience a social and conceptual world which is conditioned by the fact that most of its participants *can* make distinctions between the experiences denoted by the colour names. For example, the landmark colour names are often employed to signal distinctions in domains other than perceptual colour. Some instances of this might be the association of RED with danger (cf. GREEN), left-wing (cf. BLUE), stop (cf. GREEN), hot (cf. BLUE), and debt (cf. BLACK), or the associations of BLUE with right-wing, cold, erotic, male (cf. PINK), sky and sea, and so on. It is possible that these sorts of relations influence the subjective similarity of the colour terms for both normal and colour-blind subjects, and that it is on the basis of these more abstract relations that semantic judgements are made. Supportive of this view, that the representations which form the basis for semantic judgements need not be closely tied to perceptual representations, is the finding that reliable configurations are derivable for categories of terms which do not have direct perceptual concomitants (Burton 1972; D'Andrade, Quinn, Nerlove and Romney 1972; Rapoport and Fillenbaum 1972; Rosenberg and Sedlak 1972; Wish, Deutsch and Biener 1972).

Semantic similarity judgements, which are the data on which the MDS studies cited above were based, predict behaviour in other tasks (see Chapter 1; Young and Routh, unpublished observations), such as the degree of 'Stroop' interference in colour naming reaction-time tasks (Dyer 1973; Stroop 1935). Stroop tasks consist of the presentation of stimuli which typically comprise two features: a hue and a distractor. The hue, carried by different colour inks, must be named by the subject, while the word which is written in the ink is ignored. Interference is greatest when the hue (the ink whose colour has to be named) is a colour whose name is judged to be similar to the distractor. For example, the word RED written in orange ink gives rise to greater interference than does the word GREEN written in orange ink.

There is telling evidence, both behavioural and electrophysiological, that the locus of Stroop interference is within a semantically organized processor which is relatively late in the sequence of processes which mediate this task (Duncan-Johnson and Kopell 1981; Tominaga, Kuda, Yogi, Miyara and Tomori 1989; Seymour 1977; Stirling 1979; Warren 1972). Since the pattern of interference reflects the pattern of similarity judgements, the semantic judgement data may capture some organizational aspect of the processes at the locus of interference.

The discussion above enumerates three features of semantic similarity judgements of the colour names, and of the MDS configurations which describe them. These are the differences between the semantic and perceptual configurations, the similarity of the configurations from colour-blind and normal subjects, and the fact that judgement data predict interference in a task where the locus of interference is thought to be late. An interpretation of these aspects is that the judgements are based on relational comparison of representations which are not simply perceptual (Miller and Johnson-Laird 1976). To be sure, the extension and reference of the colour names may rest on the perceptual experience of colour, but the regularities which the underlying representations capture are unlikely to be exclusively confined to simple perceptual qualities (see also Fillenbaum and Rapoport 1971; Shepard 1972; 1980).

Most electrophysiological investigation of the colour names has been confined to studies in which Stroop stimuli have been employed (e.g. Aine and Harter 1986; Duncan-Johnson and Kopell 1981; Tominaga et al 1989). The fact that colour names were used in these studies was incidental to interests in selective attention (e.g Aine and Harter 1986) or in the effects on P300 latency of the congruence and incongruence of Stroop stimuli (e.g. Duncan-Johnson and Kopell 1981; Tominaga et al 1989). Before beginning the work reported here, however, I undertook an experiment in which the relations between the colour names were the focus of my interest. This experiment was therefore a direct precursor of the present studies.

The experiment involved recording ERPs to each of the eleven colour names selected by Berlin and Kay (1969), and a data reduction and description procedure similar to that detailed in Chapter 2. Several aspects of the data *analysis* technique described in Chapter 2, however, had yet to be developed. In particular, the employment of PROCRUSTES rotation to assess the configurational similarity of the models and the subsequent randomisation test were not undertaken. Nonetheless, a brief exposition of it is included to show the empirical base from which this work began.

Sixteen electrodes referred to a balanced noncephalic pair (Stephenson and Gibbs 1951) were placed in a montage chosen from the 10-20 system (Jasper 1958). This was designed to cover as much as possible of the scalp. Signals from the scalp electrodes, and from bipolar EOG electrodes, were amplified with a bandpass of 0.03-70 Hz, and were stored on digital tape and averaged off line. One subject (the experimenter) took part, and was passively exposed to a random sequence of the 11 colour names of Berlin and Kay (1969) (RED, ORANGE, YELLOW, GREEN, BLUE, PURPLE, GREY, BLACK, WHITE, BROWN, PINK), which were presented at an ISI of 5 seconds. Each word was presented 40 times. The sequence was broken every 3 minutes to revive the subject: a procedure necessary since the passive exposure paradigm was extremely tedious.

Figure 3.1 shows a plot of the ERP model derived on the basis of the whole post-stimulus epoch (1024 msec). Only the two most important dimensions, of a three dimensional model, are plotted, with the third dimension embedded in the plane of the paper. Figure 3.2 shows a similar plot of a model derived from 30 subjects' judgements of the semantic similarity of the target words. Despite the fact that the two models have not been rotated in to maximal convergence, both models show a polarity between BLACK and WHITE, and a somewhat circular structure having the order RED, ORANGE, YELLOW, GREEN, BLUE, PURPLE, RED: the so-called colour circle. The degree of configurational similarity between the two models was assessed, in a crude way, by regressing one proximity matrix on the other (Note that the entries in a proximity matrix are not independent measurements) and the regression coefficient was significant at $p < 0.0025$ (cf., Manly 1986; Mantel 1967).

This result (which was the expected one (see Chapter 8)) was construed as evidence that a semantic factor underlying the colour names was the major predictor of variance between ERPs evoked by passive exposure to these words. That is, the more similar two words were judged to be in meaning, the more similar were the ERPs evoked by exposure to them. The criticisms made in Chapter 1 of the studies of Chapman et al (1977; 1978), however, could also be made of this pilot study. Only one possible linguistic factor was compared to the

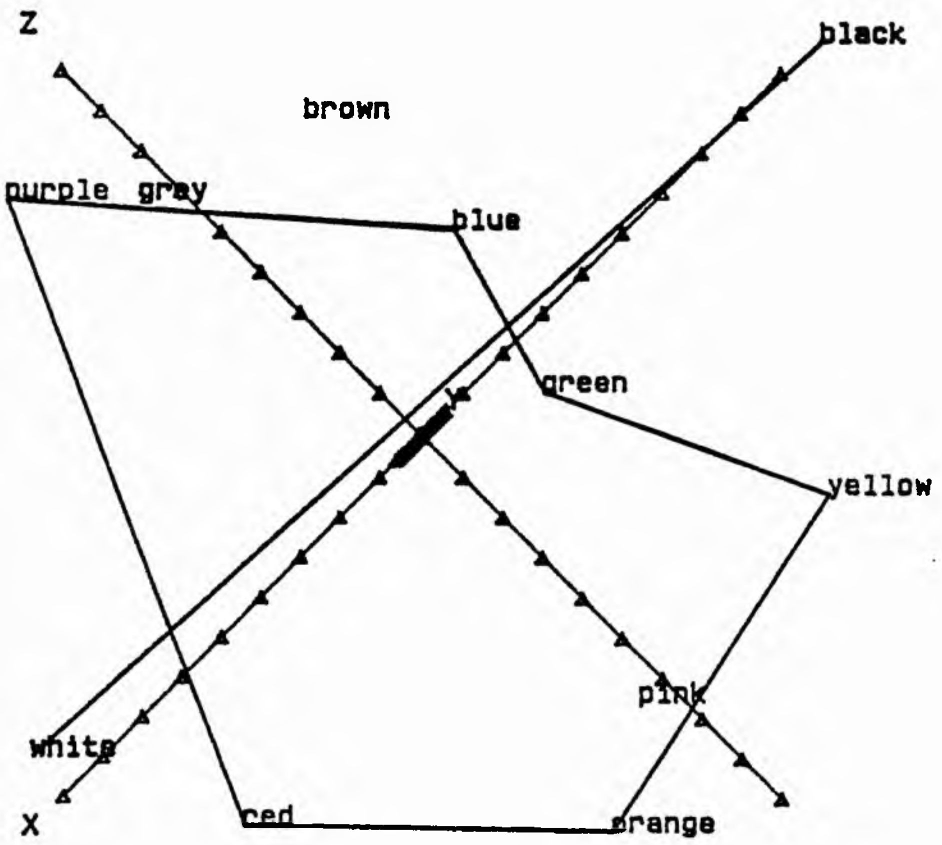

Z
brown
black
purple
grey
blue
green
Y
yellow
pink
white
X
red
orange

FIGURE 3.1

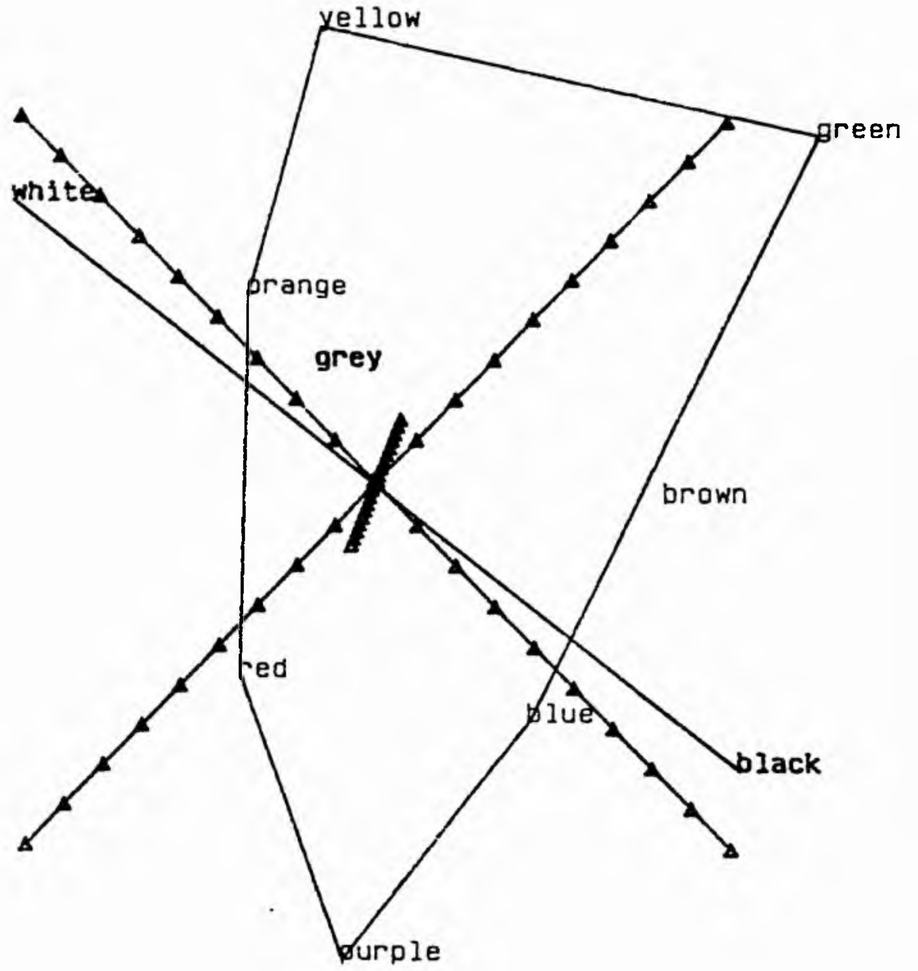


FIGURE 3.2

ERP variance model, excluding from consideration the possibility that some other linguistic factor might better explain variance between the ERPs. Nonetheless, the apparently good correspondence between the physiological and judgements models suggested that it might, indeed, be possible to explore relations between the linguistic and ERP domains in this way.

In order to more systematically explore such relations, it is necessary, as detailed in Chapter 1, to decide on a selection of word items, and to derive numerical models which represent the ways in which the selected items differ in the respect defined by each of the linguistic factors.

The choice of the number of colour words to present and analyse was conditioned by two considerations. The number of items should not be too small, in order to avoid problems of metric indeterminacy (see Chapter 2). But the number of items should not be too large, in order to avoid the problems associated with over-long experiments (see Chapter 2, in relation to multiple regression). The six colour terms discussed by Miller and Johnson-Laird (1976) were considered too small a set to analyse. The eleven colour words of the Berlin and Kay (1969) selection of 'basic colour terms', were chosen as an initial compromise between the above two considerations. Thus the words chosen for study were RED, ORANGE, YELLOW, GREEN, BLUE, PURPLE, GREY, BLACK, WHITE, BROWN and PINK; the same set, in fact, as explored in the pilot experiment. I was mindful, however, of the objections raised to some of these words with respect to their candidature as basic colour terms (Miller and Johnson-Laird 1976). The analyses in the experiments which follow are therefore based on a set of eight colour names, which is the above set minus PINK (alleged to be light red, but judgements of the relation of this term to the other colour names appeared to be erratic), BROWN (judgements of which again appeared to be erratic, perhaps due to its 'aspecificity' (Miller and Johnson-Laird 1976)) and ORANGE (which being more commonly a noun than a colour name, I felt might invoke spurious syntactic effects (Molfese 1983)).

Having selected the items, the next issue is that of finding an appropriate set of linguistic factors. Again, there are two considerations. The first is that the set of linguistic factors should be large enough to make it likely that the important factors are among the set of factors against which the ERP model will be rotated (i.e. preferably an exhaustive sample of the factors which could relate to differences between the words). The second is that, because the number of linguistic models is a multiplicative factor in the amount of computation each analysis requires, the set of linguistic factors should not be redundantly large. I was guided in the choice of an optimal set of linguistic factors by the set included by Whaley (1978).

Whaley (1978) examined the effects of 16 linguistic factors on reaction time to word stimuli in a word-nonword decision task. The factors divided into 5 groups: word length variables, letter frequency variables, digram frequency variables, word frequency variables and a group of variables from the Paivio, Yuille and Madigan (1968) norms. The latter variables; Imagery, Concreteness, Meaningfulness and Age-of-acquisition, are difficult to apply to the colour names. For example, in a number of pilot experiments, all of the Berlin and Kay

(1969) colour names were found to give rise to mental images with about the same ease. Similarly the colour names refer to about equally concrete or abstract concepts.

Age-of-acquisition (AOA) is generally operationalised by collecting subjects' estimates of the age at which they acquired use of concrete nouns (Whaley 1978). The employment of this variable here is complicated by two considerations. The first is that the colour names are not nouns in a straightforward sense (Miller and Johnson-Laird 1976). Notwithstanding Senator McCarthy, one cannot say "There is a red". Nor are they concrete in the original sense used by the discoverers of the salience of this variable in picture naming (Carroll and White 1973). The second is that the pattern of acquisition of the use of these terms has been investigated and found to be very irregular (Bartlett, personal communication cited in Miller and Johnson-Laird 1976; Bateman 1915; Hopmann 1972). Children learn the application of the terms in such a way that the use of a particular term may be reliable at one time, and completely unreliable at some point later. Moreover, the sequence in which the terms fall in and out of reliable use is highly variable between children (Miller and Johnson-Laird 1976). Quite what linguistic factor would be captured by collecting subjects' estimates of the age that they acquired the use of these particular terms seemed to me uncertain.

It is possible that estimates of these words' AOA are related to an underlying access to their subjective familiarity (see Gernsbacher 1984; Miller and Johnson-Laird 1976; Rubin 1980). Subjective familiarity is generally operationalised by collecting subjects' ratings of the familiarity of words. Unfortunately, data related to this variable, which is closely associated with word frequency (Gernsbacher 1984; Rubin 1980), were very difficult to collect. Naive subjects (but not psychologically sophisticated ones) reported that they found the request to rate the familiarity of the 11 colour names meaningless, and in most cases all the words were judged to be completely familiar.

The linguistic factors chosen by Whaley (1978) did not include factors related directly to the orthographic or phonological properties of words (although the inter-letter probability structure captured by digram frequencies might covary with orthographic regularity). While there is no generally agreed means of assessing the orthography of single words, nor the orthographic similarity of word pairs, it would be surprising if some orthographic variable were not involved in word recognition. One approach to trying to quantify the orthographic similarity of the colour words is simply to present pairs of words and to collect subjects' judgements of their orthographic similarity. A quantitative model of the orthographic structure of the colour words can then be generated by non-metric multidimensional scaling, in the same way as is standard for judgements of semantic similarity (Fillenbaum and Rapoport 1971). A similar procedure can be effected for judgements of the phonological similarity of the colour names.

The above considerations led to the choice of the following 14 linguistic factors. Word length variables are represented by 'Number of letters', 'Number of syllables' and 'Number of phonemes'. The two former variables were derived simply by counting the letters and syllables (phonologically defined), respectively, and the latter was derived by segmenting the colour names into phonemes according to a procedure suggested by Glucksberg and Danks (1975), and then counting the phonemes. Letter frequency variables were derived from Baddeley, Conrad and Thompson (1960), and are 'First letter frequency', 'Last letter frequency' and 'Mean letter frequency'. The digram characteristics of the words, also derived from Baddeley et al (1960), are signalled by 'Mean digram frequency'. The Baddeley et al (1960) data were not word length specific. Word frequency variables are 'Kucera-Francis I' (gross number of counts in the Kucera and Francis (1967) corpus), 'Kucera-Francis II' (a measure of dispersion in language), 'Kucera-Francis III' (which shares properties of both Kucera-Francis I and Kucera-Francis II), and 'Thorndike-Lorge' (from Thorndike and Lorge 1944). The semantic relations of the colour names are represented by an MDS model based on subjects' judgements of their semantic similarity. The orthographic relations of the words are captured by an MDS model based on subjects' judgements of the words' orthographic similarity, and the phonological aspects by a similar model derived from judgements of phonological similarity.

This cannot be claimed to be an exhaustive sample of the possible linguistic factors. New factors may be discovered for which no framework exists at present, and only a subset of the factors noted in Rubin (1980) can be meaningfully related to a relatively homogeneous set of words such as the colour names. For example, a number of the variables employed by Rubin (1980) relate to a variety of timed tasks (e.g. Lexical decision, naming latency, free recall with overt rehearsal and so on). These variables seem better thought of as behavioural dependent variables rather than as linguistic factors in their own right. Similarly, some of the factors relate only to concrete nouns, or are otherwise inapplicable to the domain of colour names, as discussed above. Nonetheless, the selection of linguistic factors is at least as inclusive as that of Whaley (1978), due to the replacement of the meaning-related variables of Paivio et al (1968) by the semantic similarity model, and the inclusion of models intended to relate to the orthographic and phonological structure of the colour words. Also, for the linguistic variables chosen (with the possible exception of 'number of syllables') there is sufficient variability in the models that variance in the ERP models due to any of the factors should be identifiable (see Whaley 1978). The actual numerical values of the linguistic models of the colour names are attached in Appendix 2. The logical aspects of the conclusions which can be drawn from the relation of this set of factors to ERP variance are discussed in Chapters 5 and 8.

Experiment 1

Introduction

In the design of this first experiment I was at pains to ensure that as little as possible extraneous variance, such as that due to response production, intruded into the waveforms of the ERPs evoked by each of the colour words. Accordingly a paradigm intended to be somewhat similar to that of Kutas and Hillyard (1980a) was chosen. This involved the subjects in attending the word stimuli but did not require a response during the presentation. Kutas and Hillyard (1980a) presented sentences and instructed the subjects to attend in order to answer questions on their content after the experiment. In this case, the paradigm entails the presentation of single word stimuli, and the same instruction was not applicable. The paradigm simply involved the attended, but passive, exposure to verbal material.

Method

Subjects: Six young adults were payed for participation. Each subject filled in a questionnaire from which the following features are drawn. Four subjects were female. All but one reported themselves to be right handed. All subjects learnt English as their first language. All were naive as to the purpose of the experiment. None was colourblind and none was being prescribed medication at the time of the experiment. No subjects had taken part in an evoked potential study previously.

Design and task: Each subject took part in four sub-experiments which together comprised experiment 1. These were an ERP recording session, and three judgement tasks during which EEG was not recorded. The order of these tasks was randomized (Note that the 24 possible orders were not systematically sampled: the design was not fully counterbalanced in this respect), in such a way that half the subjects performed a judgement task before the ERP recording session. All four tasks were undertaken in a single two hour experimental session.

ERP design and task: The ERP experiment consisted of the passive exposure of each subject to a pseudo-random sequence of the 11 colour names, and averaging an ERP from the 50 presentations of each word. The subjects were instructed simply to attend to each word and to try to blink only in the inter-trial interval, during which a fixation asterisk was displayed. In factorial terms, the design was simply a word condition factor (with 11 levels) by an electrode channel factor (with 13 levels) with repeated measures.

ERP Recording:

Electrodes: Burden Ag\AgCl electrodes were fixed with collodion at 13 10-20 system (Jasper 1958) sites; FPz, Fz, Cz, Pz, Oz, F7, F8, T3, C3, C4, T4, T5 and T6, and were referred to a balanced clavicular-vertebral pair of electrodes (Stephenson and Gibbs 1951). Ground was placed between Cz and Pz. EOG was recorded bipolarly from the outer canthus of the left and from above the right eye, by silver cup electrodes attached with a double-sided adhesive ring. Each electrode site (except for the EOG electrodes) was first degreased by wiping with cotton wool soaked in alcohol. The skin beneath all electrodes was prepared by agitating it with a 'Q-tip' soaked in 'Omniprep'. In most cases the electrode impedances were immediately less than 5 KOhms, but in those cases that were not, the skin was abraded through the hole in the centre of the electrode with a plastic pipette tip. Electrode impedances were, after this procedure, always less than 5 KOhm.

Amplification: All signals were amplified by SLE amplifiers (40 uV/V for the EEG electrodes and 320 uV/V for the EOG) set to a time base of 5 seconds and a high pass of 32 Hz (3dB down). This bandpass is that typically employed in this laboratory (e.g. Rugg 1987) and represents the nearest to D.C. recording of which the amplifiers are capable. The high pass was intended not to be so low as to reduce the amplitude of components likely to be recorded in this paradigm, but to attenuate high frequency muscular artifact (Electromyogram: EMG).

Recording and digitisation: The signals were digitized at a rate of 4 msec per point by a Cambridge Electronic Design CED 1401 laboratory interface at 12 bit resolution. The sensitivity of the digitisation was equivalent to 0.1 uV per digitisation increment for the EEG channels, and to 0.8 uV per digitisation increment for the EOG channel. Each trial's data were immediately written to computer hard disc.

The recording epoch began 100 msec prior to stimulus onset and continued for 924 msec after presentation.

Stimulus presentation: The inter-stimulus interval was 5 seconds, during which a fixation asterisk was displayed. Each stimulus word was presented in capital letters on a TV monitor about 1 metre from the subject. Each was on screen for 292 msec and subtended approximately 2 by 0.5 degrees of visual angle. The letters, of which the words were composed, were generated by a 5 x 7 white dot matrix on a dark (video off) background, and each stimulus was presented at well above threshold contrast luminance.

The sequence of events comprising one trial is summarised in Figure 3.3.

ERP averaging: Averaging was performed off-line by a locally written computer program. Differences of D.C. offset in each amplifier channel were removed from each trial's data by subtraction of the average of the 100 msec prestimulus baseline. Trials in which eye-movement artifacts did not give rise to signals which exceeded a preset threshold, and which contained no extreme

Events comprising one trial of experiment 1

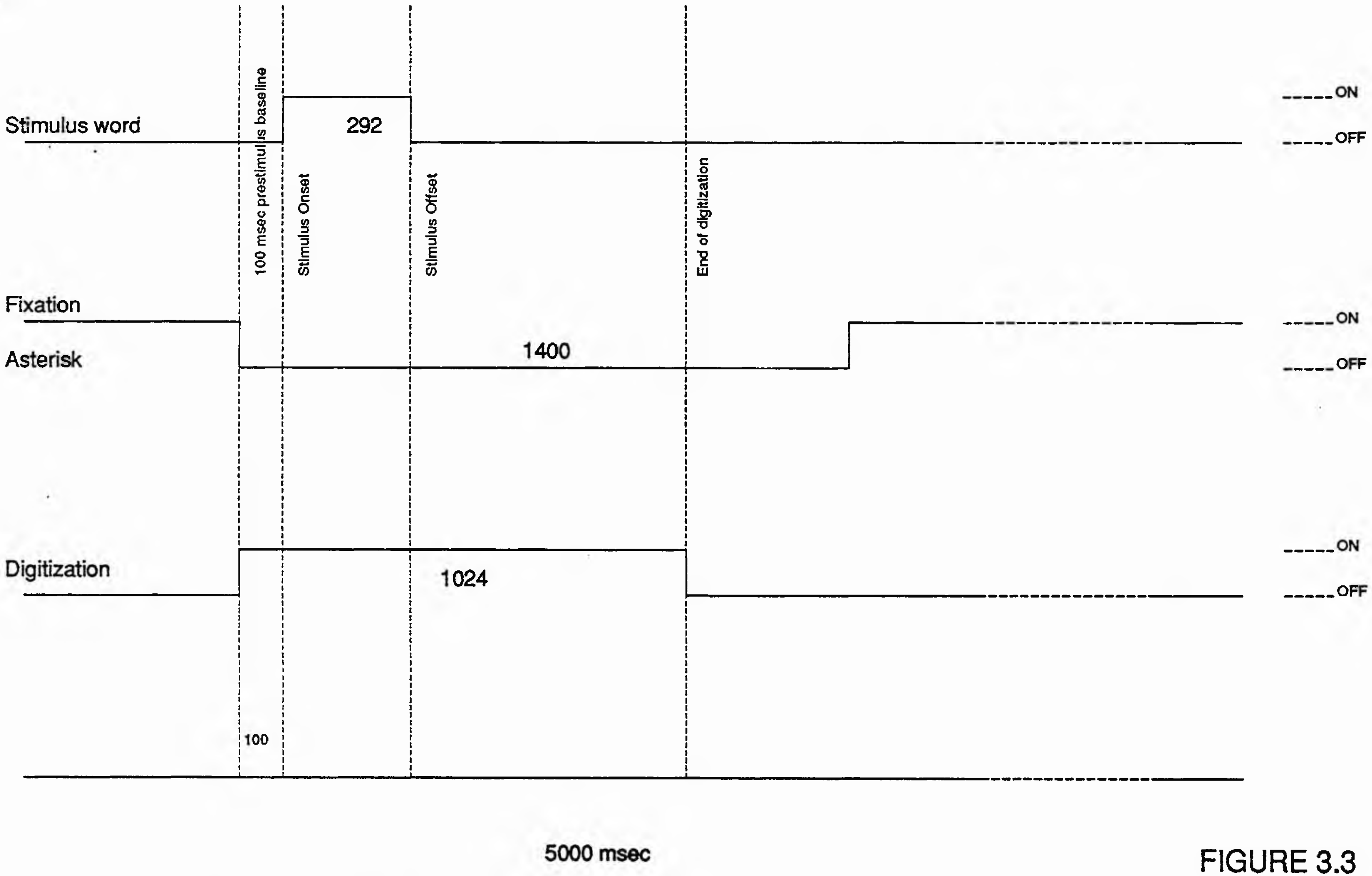


FIGURE 3.3

values or amplifier clipping on any channel (such as might result from gross movements of the subject) were summed and then divided down by the number of trials comprising the sum in the relevant condition (i.e. the 'bin' number). The bin numbers were always appended to the header information in the averaged ERP data file which was then written to hard disc.

ERP experiment procedure: After application and impedance testing of the electrodes and, in appropriate cases, performing one or more of the judgement tasks, subjects were seated in a soft upright chair. The clavicular-vertebral reference electrode pair were balanced by adjusting a potentiometer to remove heart artifact from the EEG trace. Facing the subject was a table on which was the TV monitor. The task was explained, as above. Subjects were also asked to try to withold blinking during each trial. It was ensured that the subjects were comfortable, understood the task and had no problem seeing the words. With respect to the latter aspect, subjects had the option of altering the ambient lighting level. When the subject was ready, the experiment began. The presentation was composed of 5 blocks of 100 trials and one of 50 trials, between each of which there was a short break. Subjects were debriefed during removal of the electrodes.

Judgement Task design: In addition to the ERP recording session, subjects performed three similarity judgement tasks. These involved pairs of colour words being presented on a computer screen. Subjects rated the similarity of the presented pairs under three sets of instructions. These were designed to collect judgements of the orthographic, phonological and semantic similarity of the colour names. The resulting proximity data were submitted to 3-way nonmetric multidimensional scaling (Carroll and Chang 1970) to generate metric models to express these three ways in which the words differ.

Judgement procedure: Subjects were seated in front of a computer, on the screen of which the first of a series of practice trials was displayed. This display consisted of a pair of words (e.g. RED vs WHITE), in white letters on a dark (video off) background, below which was a scale marked from 1 (very similar) to 9 (very different). The subject was instructed as follows. For the semantic similarity judgement task, the instructions were to "judge the similarity in meaning of the two presented words and assign a value from the scale below. Try to make your judgement on the basis of the similarity in meaning of the words, quite apart from what the presented pair of words may look or sound like". The subject then entered the value of their judgement of the semantic similarity of the two words.

On doing so, the next pair of words, which were pseudo-randomly selected, were displayed and the process of judgement repeated. If an error occurred (such as entering a number not between 1 and 9), the pair of words was repeated at a randomly chosen point later in the presentation sequence. Judgements were always self paced. The sequence was repeated for a large enough number of practice trials to ensure that the subject understood the task, how to operate the computer and had seen all the stimulus words.

The computer presentation program was then interrupted and restarted with a different random number seed, so that the sequence of word pairs would not be the same as for the practice trials. After completing n(n-1)/2 = 55 judgements (i.e. all-pairwise comparisons sufficient to fill a lower triangular proximity matrix), the matrix was written to disc with header information showing the subject's identification and the task to which the judgements related. Previous work (Young and Routh, unpublished observations) showed that judgements were unaffected by the order in which the two colour words were presented. The collection of a full proximity matrix (i.e. not just the lower triangular part) was therefore judged unnecessary, particularly since the collection of a full matrix would have required the judgement tasks to have taken twice the time to perform.

The orthographic similarity judgement task was operationalised by instructing the subject to "judge how similar the two presented words look and assign a value from the scale below. Try to make your judgement on the basis of the similarity in appearance of the words, quite apart from what the presented pair of words may mean or sound like". As above, for the semantic similarity judgement task, the procedure involved practice trials, restarting the computer program with a different random number seed and writing the judgement data to disc. The same procedure was undertaken for the phonological judgement task, save that the intruction was to "judge how similar the two presented words would sound and assign a value from the scale below. Try to make your judgement on the basis of the similarity of their sounds, quite apart from what the presented pair of words may mean or look like".

Both the ERP data and the behavioural judgement data were transmitted to a VAX 11-785 mainframe computer for analysis.

Results and Discussion

The results of this experiment are complex, and reporting them is aided by some associated discussions of this first worked example of the data reduction and analysis procedures.

Behavioural data: The semantic similarity judgement data were submitted to a non-metric multidimensional scaling program (INDSCAL, Carroll and Chang 1970). As described above, all analyses were based on a set of 8 colour names. In order to ensure that the 3 judgement models reliably reflected the aspect of the words they were designed to capture, it was decided to employ judgement data from this experiment and the two which follow (n = 18). The reason for this was that judgement data derived from the 3 tasks were were almost identical across the 3 experiments. Models based on data from all 3 experiments were therefore selected to reflect the semantic, phonological and orthographic aspects of the words, on the grounds that an N of 18 would allow very accurate estimation of the parameters of the models, and would be in keeping with the numbers of subjects typically employed in psychological scaling experiments (e.g. Rapoport and Fillenbaum 1972).

The actual routine used was one of the model options of the versatile ALSCAL program (Young, Takane and de Lewyckyj 1978) instantiated in the SPSS-X statistical package. The procedure was to present the 18 judgement matrices as input to the program, as is standard in this area (e.g. Fillenbaum and Rapoport 1971). The program then computes a 'stimulus space' which is a metric model of the configuration of points which best accounts for the variance in the 18 input matrices. The goodness of fit of the model to the input parameters is expressed as an average of the variance explained by the model in each of the proximity matrices. The program also produces a 'subject space' which shows the degree to which each subject weighted each of the computed dimensions (Carroll 1972; Wish et al 1972).

I turn first to the judgements of the semantic similarity of the colour names.

The goodness of fit of the model to the input matrices is related to the number of dimensions selected for the output. In the same way that successive principal components add to the overall variance-explained, the more parameters in the fitted MDS model, the more variance is explained. This gives rise to a problem similar to that of deciding how many PCs to retain to explain data parsimoniously but accurately: the problem of the optimal number of dimensions in which to express the interrelation between the stimulus objects.

There are a number of guides to the choice of the number of dimensions (Shepard 1972). First, if reciprocal average variance-explained is plotted against the number of dimensions of a series of solutions of increasing dimensionality, there may be an 'elbow' beyond which there are diminishing returns in adding dimensions. This is what is often called a 'scree test'. The number of dimensions above which there are diminishing returns could be selected as an appropriate one for the data in question. In practice, however, I have never seen such an 'elbow', nor was one present in solutions for these data. This means of deciding dimensionality was not particularly useful.

A second guide is that the solution should not possess so many parameters (i.e. so many dimensions) that the estimation of them is unreliable (see Chapter 2, in relation to 'metric indeterminacy'; Shepard 1972). But this is not helpful either: solutions of dimensionality more than 10 could be estimated quite reliably from the present data.

A third guide, which is useful, is that the solution should be interpretable (Shepard 1972). In most cases interpretability is dependent on a solution of fewer than 4 dimensions, since spaces of greater than 3 dimensions are very difficult to visualize (cf., Wish et al 1972). On this basis, I chose a solution for the semantic similarity judgements of the colour names with 3 dimensions. This was to accord with the organization placed on the domain by inference from colour mixing and perceptual similarity judgements (Miller and Johnson-Laird 1976), to accord with previous MDS-based studies of semantic aspects of the colour names (Fillenbaum and Rapoport 1971; Young and Routh, unpublished observations), and because the 3-dimensional solution appears to best capture an interpretable structure.

The computed configuration in 3 dimensions explained an average (RMS R^2) of 0.725 of the variance in the judgement matrices. All 18 subjects' judgement data were well fitted by the resulting model.

The configuration of points of the computed 'stimulus space' is shown in Figures 3.4a and 3.4b. Figure 3.4a depicts dimension 1 (horizontal) plotted against dimension 2 (vertical). It can be seen that there is a polarity between BLACK and WHITE, midway between which is GREY. Distributed in roughly circular fashion are the colours RED, YELLOW, GREEN, BLUE and PURPLE: the colour circle. As remarked in the general introduction to this chapter, the polarity between BLACK and WHITE is also reflected in a polarity between YELLOW and BLUE. These latter two colours are, respectively, close to WHITE and BLACK, possibly reflecting the fact that the prototypical representations of these colours are relatively light and dark. RED and GREEN are about midway on the light-dark dimension.

Figure 3.4b shows dimension 1 (horizontal) plotted against dimension 3 (vertical). Dimension 3 seems to relate to a distinction between chromatic and achromatic colour names, since all the chromatic colours are concentrated in the top half of the plot, and all the achromatic colours are concentrated in the lower half. Subjects evidently tended to judge words in the chromatic and achromatic colour name groups as more different between these groups than within a group.

Structure derived from semantic similarity judgements of the colour names

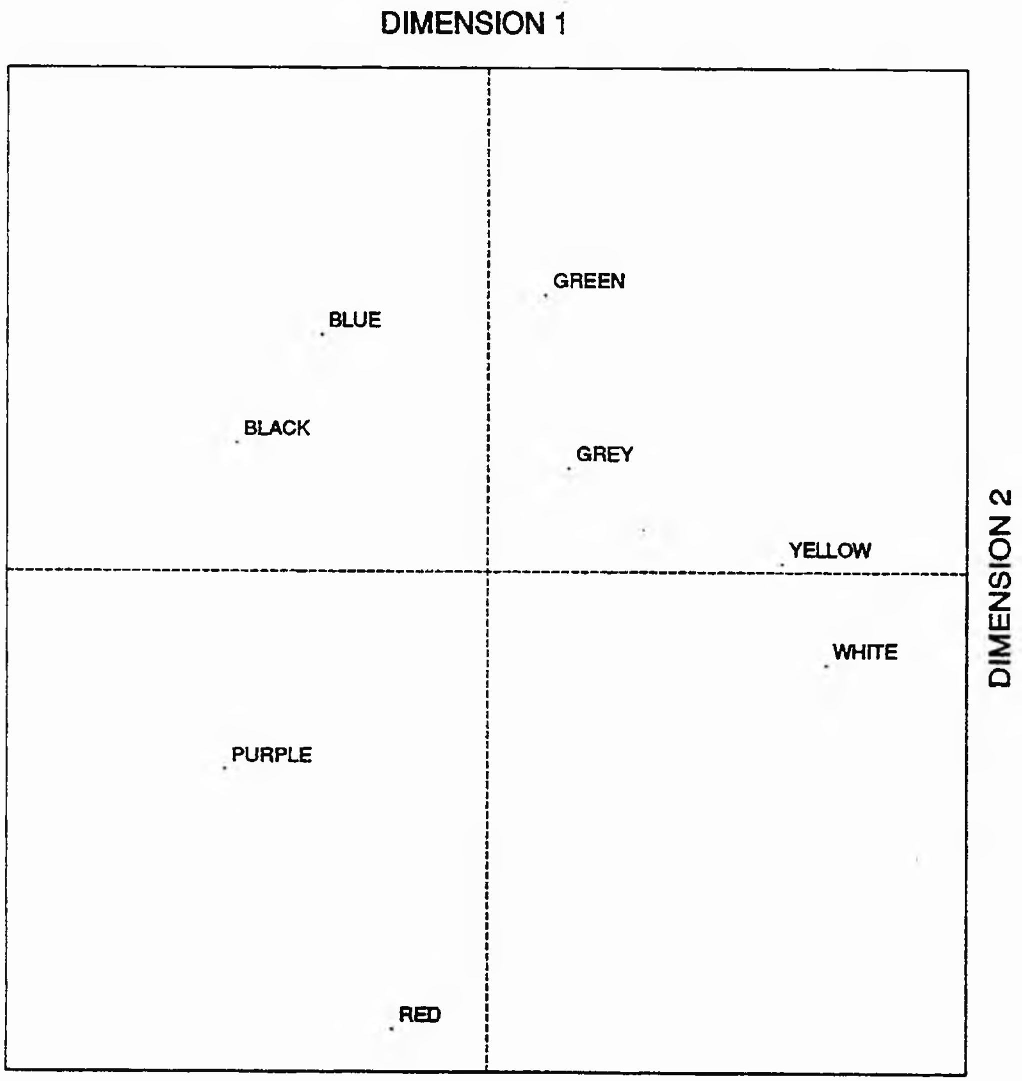


FIGURE 3.4a

Structure derived from semantic similarity judgements of the colour names

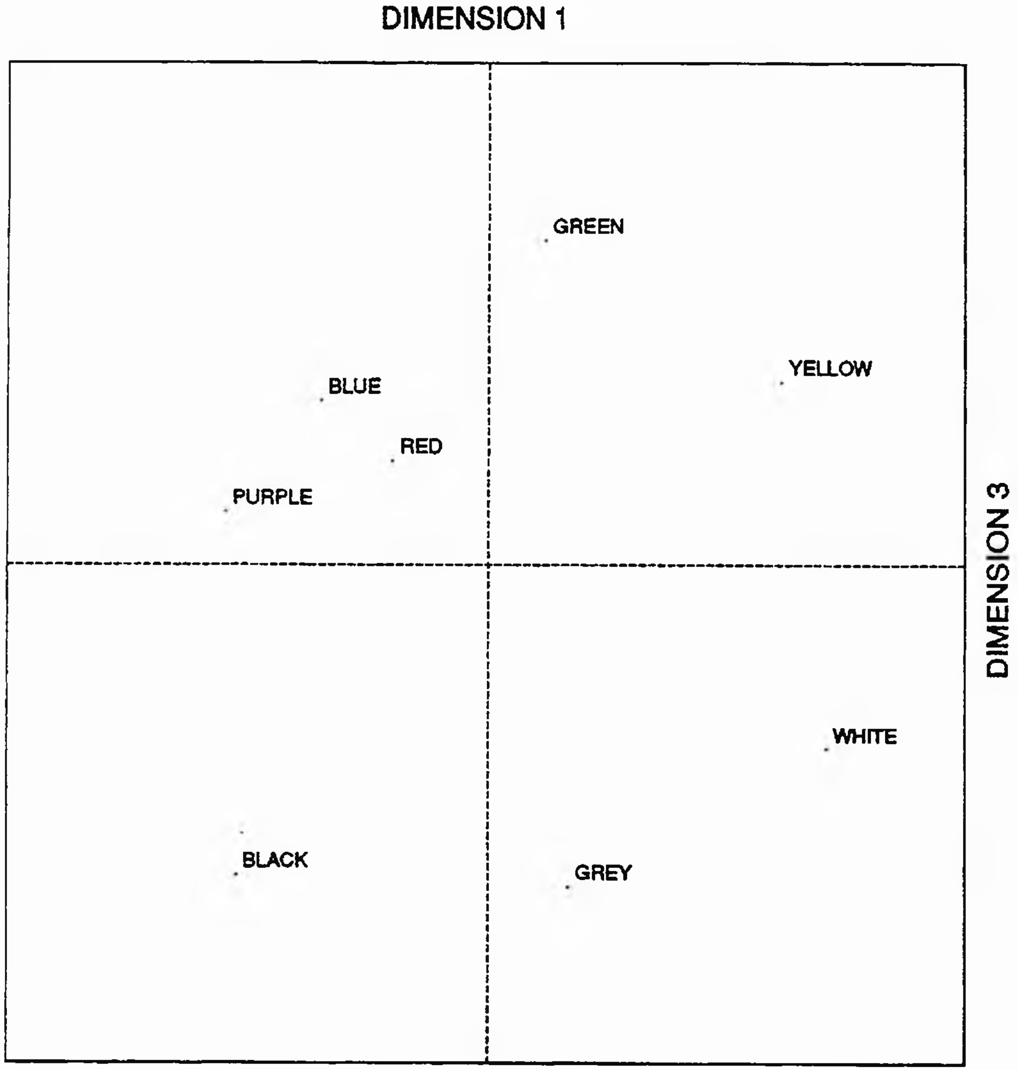


FIGURE 3.4b

The results of the analysis of the semantic similarity judgement data were in accord with previous work (Fillenbaum and Rapoport 1971; Young and Routh, unpublished observations), save that the apparent distinction between judgements of terms denoting chromatic and achromatic colours has not been noted. The coordinates in the 3 dimensions of the stimulus configuration were abstracted from the output and used as the semantic judgement model against which to compare the model derived from the ERPs. The actual numerical values of this model are attached in Appendix 2, as are all the linguistic factor models.

The orthographic similarity judgement data were submitted to an exactly analogous analysis procedure. Again, I chose a 3-dimensional solution. This was on the grounds that the 3-dimensional model appeared interpretable, and that if the intrinsic dimensionality of the words were lower than 3, the lower-dimensional structure could be embedded in 3-space with little distortion. I thought it unlikely that orthographic differences between the colour words would require more than 3 dimensions.

The INDSCAL model yielded an average variance-explained value of 0.598, reflecting greater differences between subjects in these judgements than for the semantic judgements. The relation between the colour words when their similarity was judged on an orthographic basis is shown by the configuration in Figure 3.5. As best revealed by a plot of dimension 1 (horizontal) against dimension 3 (vertical), there appear to be two tight clusters of words and 3 words

Structure derived from judgements of the orthographic similarity of the colour names

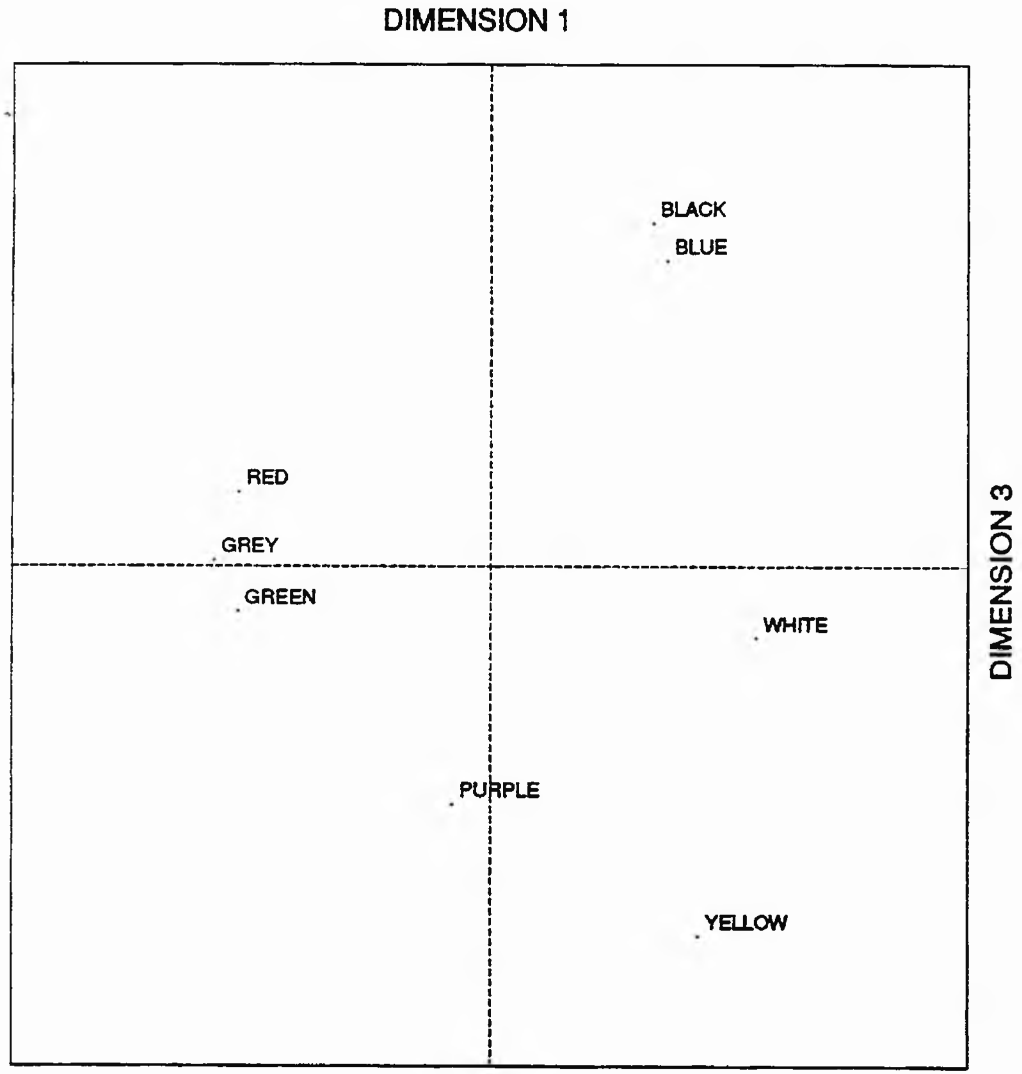


FIGURE 3.5

which are less close to other colour names. One cluster is comprised by RED, GREEN and GREY. This cluster all possess the 'RE' digram, and GREEN and GREY possess the 'GRE' trigram in common. The other cluster is that of BLACK and BLUE, which seem to be judged to be related in virtue of their possession of the 'BL' digram. YELLOW, WHITE and PURPLE are not tightly clustered with other colour words.

It is possible that there is a relation to word length in the orthographic configuration. This can be seen by imagining a number line stretched between RED and YELLOW. Mapping the position of each term onto this line, there is a roughly monotonic relationship with word length, which may or may not be related to the basis of subjects' judgements in this task. It would seem natural for subjects to judge the orthographic similarity of words on the basis of the constituent letter patterns of the words, and on their length (long and short words would presumably tend to be judged dissimilar). In any case, the configuration produced from data in this task seems to have captured an aspect of the words' differences which is interpretable as being due to their orthography. As above, the coordinates which make up the solution configuration were abstracted and employed as the orthographic model for these words (see Appendix 2).

Subjects' judgements of the phonological similarity of the colour names were again used to produce a 3-dimensional solution. The average R^2 of this solution was 0.651, a value intermediate between the values for the semantic and orthographic solutions. The output configuration is presented in Figure 3.6 which depicts dimension 1 plotted against dimension 2.

Not surprisingly, words that look similar tend also to sound similar. The solution is quite similar to that derived from orthographic judgements. BLACK and BLUE are again clustered together, as are GREEN, GREY and RED. In this solution, though, YELLOW and PURPLE are grouped together, possibly due to their both being disyllabic words ending without a hard consonant. WHITE was judged quite dissimilar to all the other colour names with respect to its sound. The coordinates of this solution were employed as the phonological model (see Appendix 2).

Interrelations between the linguistic models

The linguistic factors embodied in the colour names are not orthogonal. The interrelations between linguistic factors have presented a perennial problem for similar studies in the behavioural domain, mainly due to the application of multiple regression (see Chapter 2). One way to uncover the relations between linguistic factors is to undertake a PCA and to note the pattern of weighting of each linguistic variable on the extracted PCs (e.g. Rubin 1980; Whaley 1978). Another way, more in keeping with the general approach being taken here, is to construct a proximity matrix in which each entry reflects the amount of shared

Structure derived from judgements of the phonological similarity of the colour names

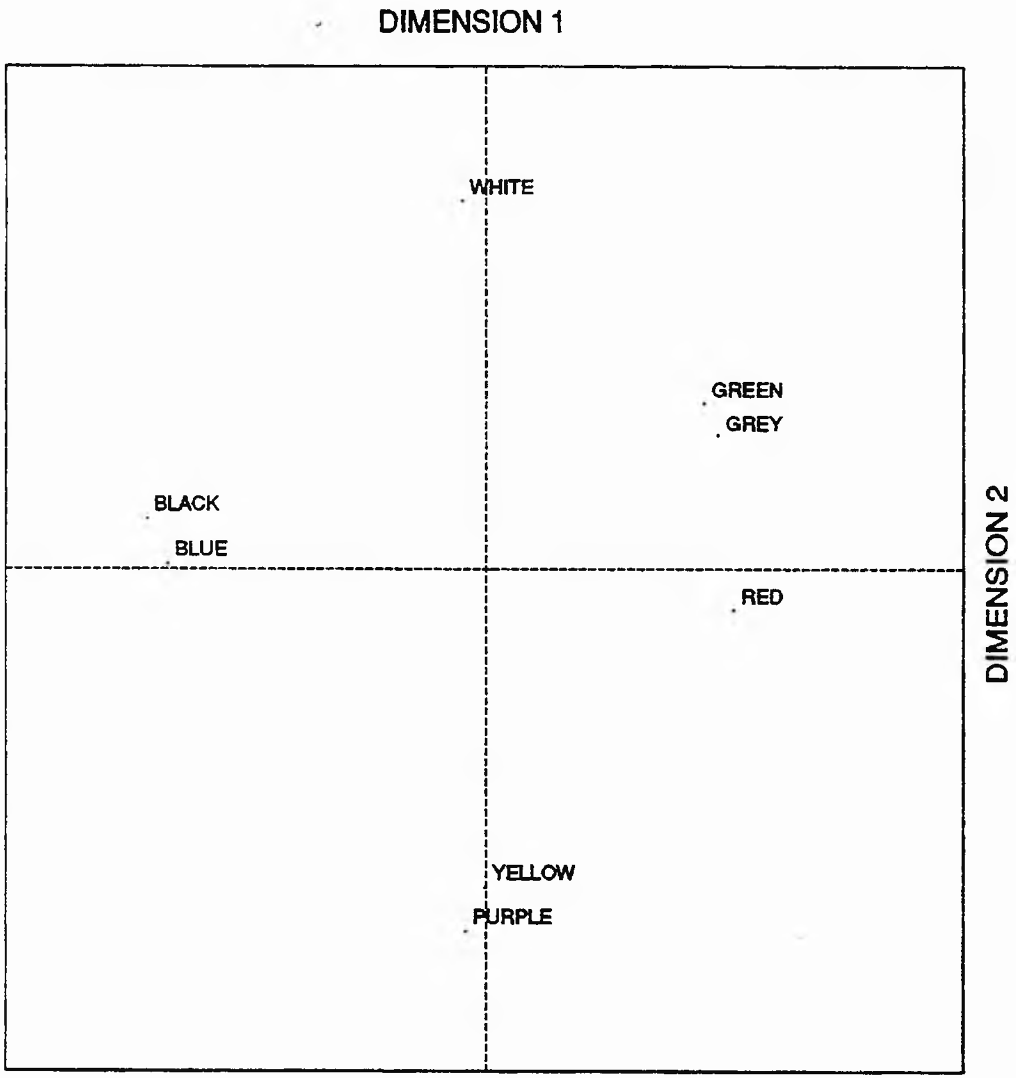


FIGURE 3.6

variance between two linguistic models. This can be accomplished by employing PROCRUSTES rotation. Rotations are performed all-pairwise so that each model is compared with every other, in order to derive variance-explained statistics to fill the proximity matrix. This takes advantage of the fact that the variance-explained statistics output by the PROCRUSTES routine are certainly a metric, and may be a euclidean metric (Alvey 1980), so that a matrix of such values can be analysed by a scaling procedure.

A matrix of proximities between the linguistic models was constructed according to this procedure. The matrix was submitted to MDS and a 2-dimensional configuration derived which shows graphically the interrelations between the linguistic models of the colour names. This is portrayed in Figure 3.7.

The first aspect to note is that the models are widely spread throughout the space, and none is in close proximity to any other. Nonetheless, regions of the space appear to be interpretable in terms of the aspect of the words that they capture. Left of the centre of the plot is a region which seems related to word frequency, since 'Kucera-Francis 1', 'Kucera-Francis 3' and 'Thorndike-Lorge' are present. Similarly the top right quadrant of the plot seems related to word length, since 'number of letters', 'number of phonemes' and 'orthography' (see above) are present. The semantic model is well dissociated from the other models, and occupies the bottom right quadrant.

Interrelations between the linguistic factor models for the colour names

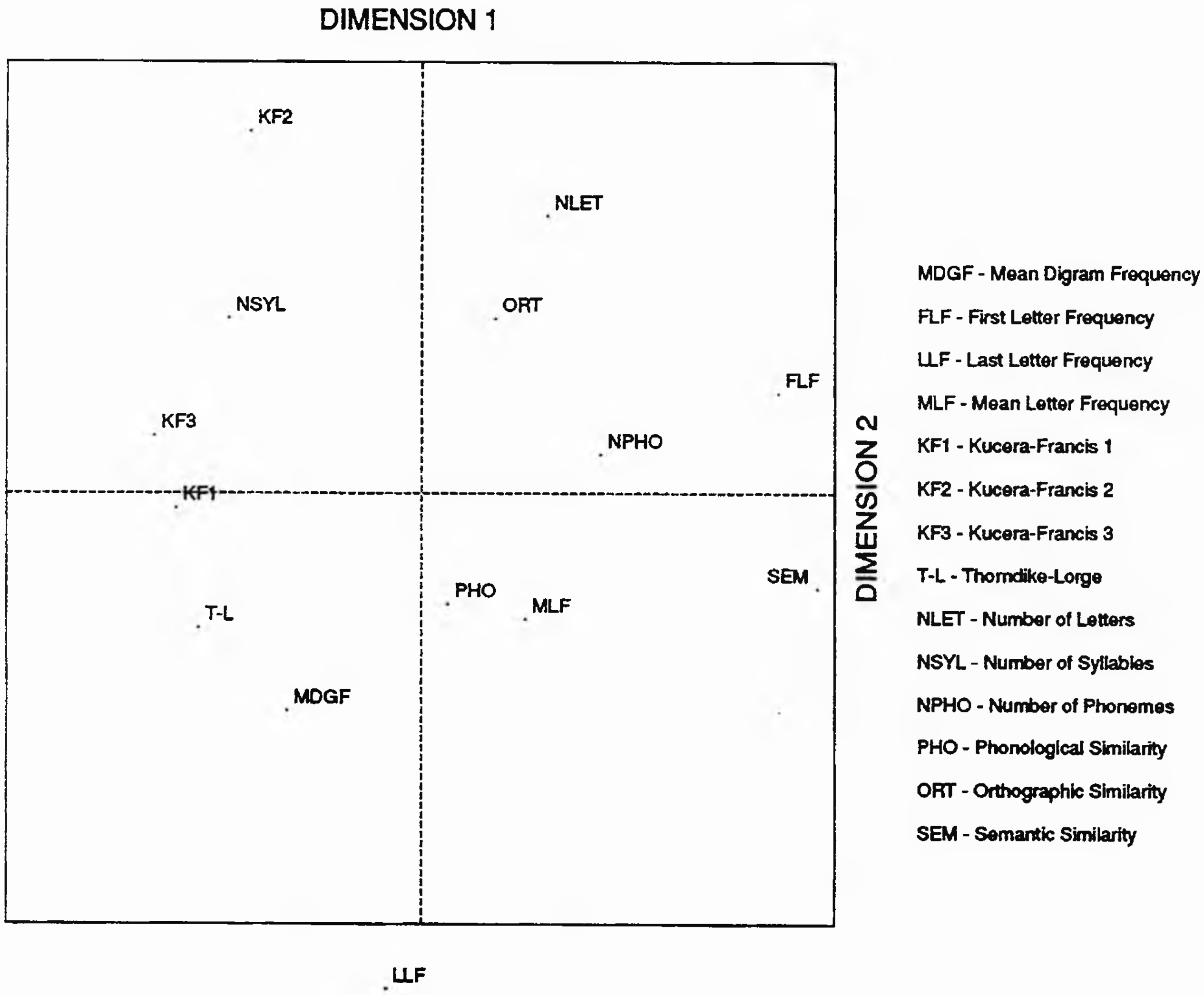


FIGURE 3.7

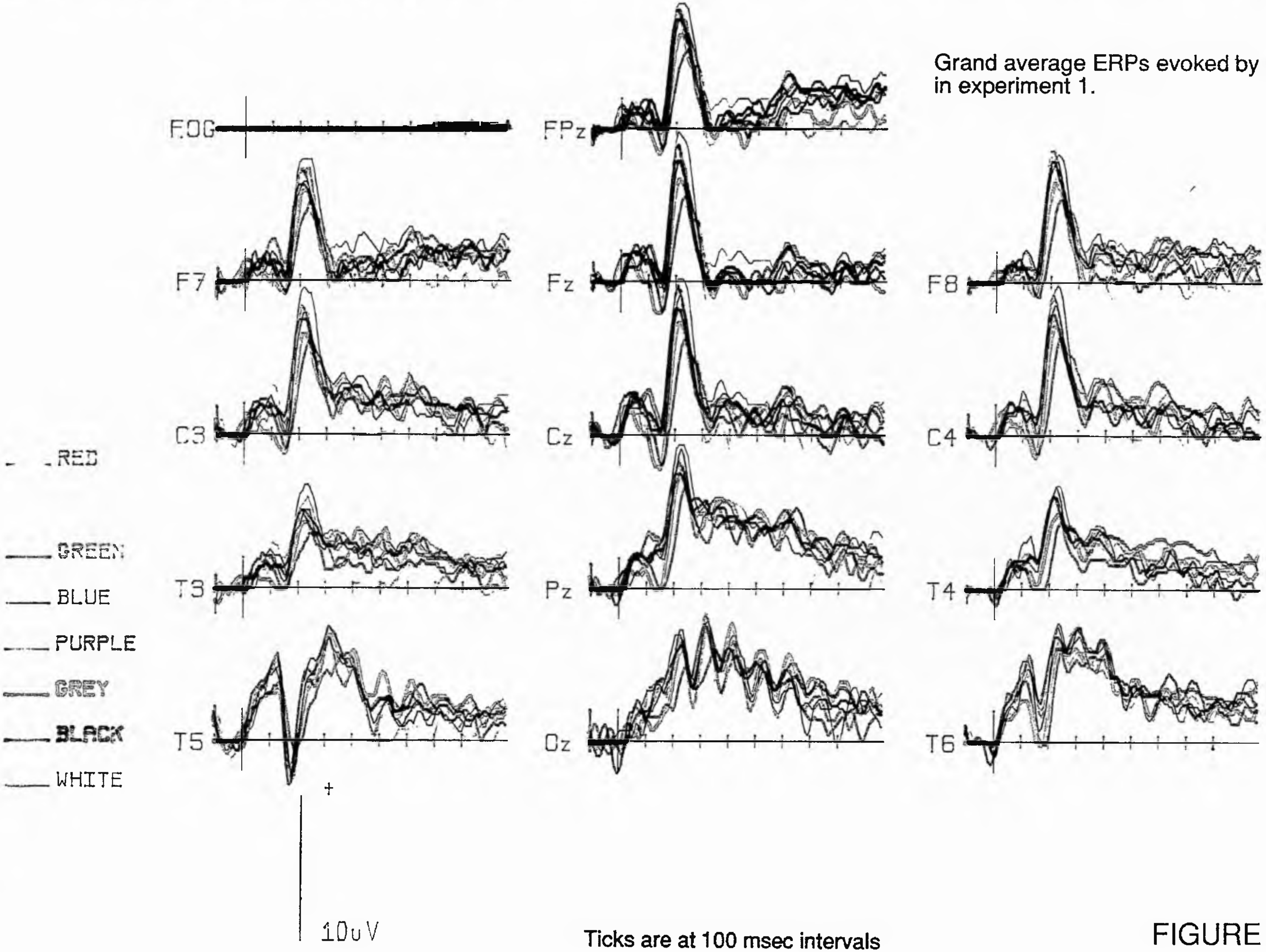
The plot in Figure 3.7 may provide a helpful way to conceptualise the relation of the ERP models to these linguistic factors. This is to conceive of the tables of variance-explained statistics associated with each ERP model as indicants of the 'gravitational attraction' on the model from each of the linguistic factors against which it is compared. Thus the ERP model could be thought of as occupying a position in the space depicted in Figure 3.7. It's position in the plane reflects the effect of the linguistic factors on ERP variance. If one particular model, or a model group, explains much more variance than competing models, the ERP model will be close to that model or region of the space.

Electrophysiological data

Figure 3.8a shows the grand average waveforms (N = 6) evoked at each electrode by each colour word. The median number of trials comprising each subject's average ERP, from which these waveforms have been derived, was 47 trials. Figure 3.8b shows, for clarity, the same waveforms, but only at the vertex electrode (Cz), and at a greater scale. In both figures, as far as possible, the colour word ERPs are plotted in the colour that each word denotes. The grand average waveforms did not enter the analysis and are presented to show the general morphology of the responses only.

The pre-stimulus baseline contains a small potential which I interpret as being due to the fixation asterisk offset and which appears to be the same for all conditions. The morphology of the post-stimulus ERP consists of a series of positive and negative deflections. First is a positive-going P1 potential which

Grand average ERPs evoked by each of 8 colour words in experiment 1.

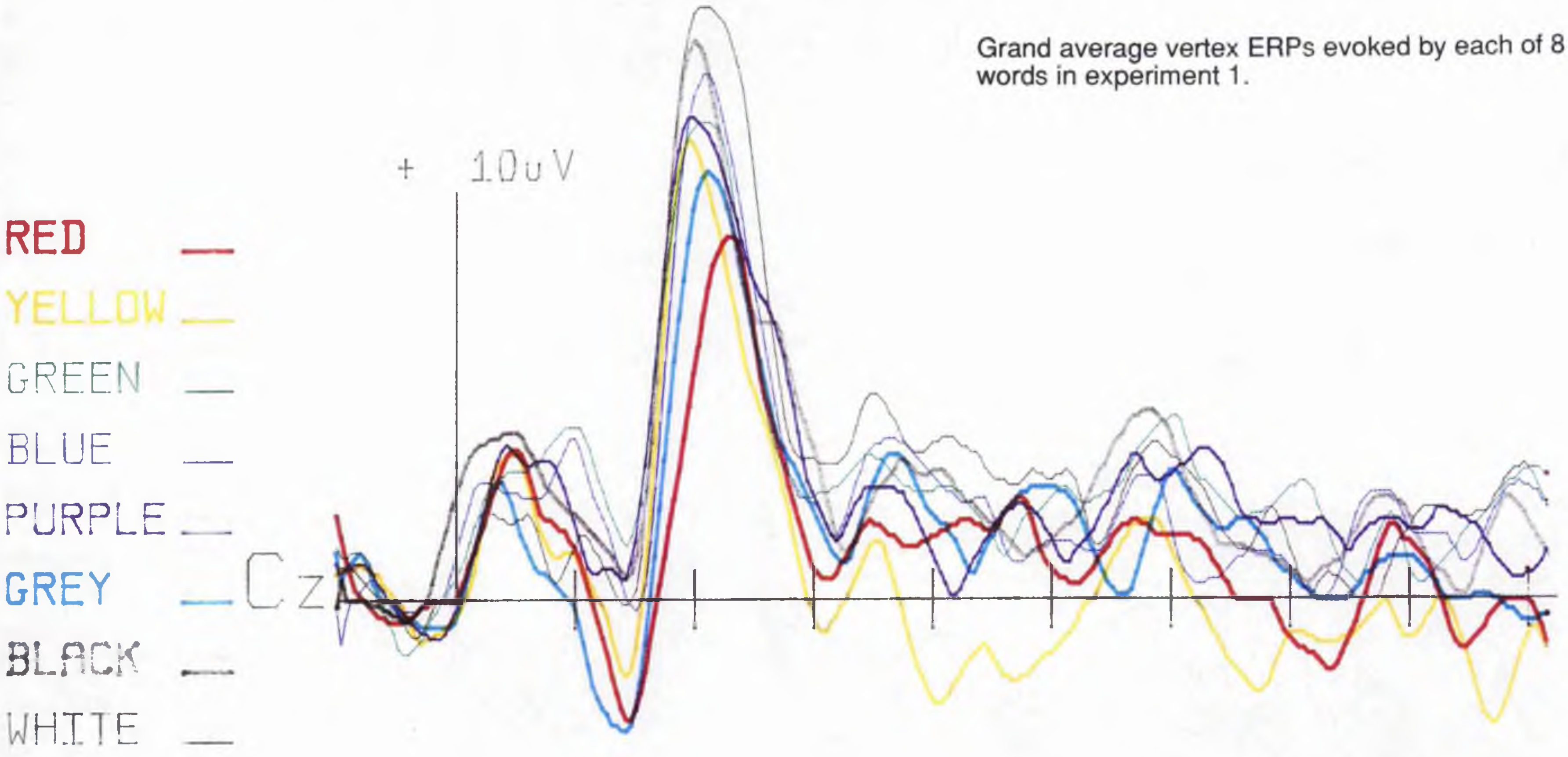


Ticks are at 100 msec intervals

FIGURE 3.8a

ERP's fro
In the
centre o
presented
subjects
handed
orthograp
nested (
(Semantic

Quantitat
words wer
and mult
in gree
compared
models f
by PROCR
explained
blue). T
was asse
test us
possible



Grand average vertex ERPs evoked by each of 8 colour words in experiment 1.

Ticks are at 100 msec intervals

FIGURE 3.8b

peaks around 70 msec. This deflection is followed by a negative potential, N1, maximal for most traces at 140 msec. This, in turn, is followed by a large P200 deflection whose peak latency is at about 210 msec. The later part of the epoch is relatively flat, and there are no late positivities, such as those related to P300.

ERP data from a time window extending for the entire post-stimulus epoch were submitted to the configurational analysis developed in Chapter 2. The first step in this process was the derivation by profile comparison and MDS of models of the variance between conditions within each subject. As with the analyses of the judgement data, a choice had to be made of the best number of dimensions for the solution configurations. There was no discernable 'elbow' in a scree test undertaken with these data, and the number of input parameters could support reliable estimates of output parameters in many dimensions. The maximum dimensionality, however, of any of the individual linguistic factor models was 3, and so if the ERP solutions were of higher dimensionality, the higher dimensions might not contribute to the desired identification of effective lingustic factors. Hence, on the grounds of parsimony and interpretability 3-dimensional solutions were chosen for the ERP models. This does not preclude the possibility that linguistic factors of lower dimensionality may be responsible for the variance in a 3-dimensional model: an essentially unidimensional model could be embedded in 3-space, in the same way that a number line can span a higher dimensional space (see below).

The average variance-explained for these within-subject analyses was 0.77. The smallest R^2 for any subject was 0.60. For comparison I undertook the same analyses on data derived from a random number generator. Of six runs the average R^2 was 0.11, and the highest R^2 of any run was 0.16. I took the difference in R^2 between the randomly generated and electrophysiological data analyses as suggestive that the within-subject ERP models may reflect some systematic differences between the waveforms (see Stenson and Knoll 1969).

The next step in the data reduction process was the derivation, again by profile comparison and MDS, of a global (across subjects) variance model (see Figure 2.1). This model explained 0.41 the variance in the input matrices. Figure 3.9 shows the two most important dimensions of the resulting ERP model. This represents the structure of differences between waveforms evoked by exposure to the colour names on the basis of all subjects: it is the global variance model depicted in relation to the rest of the data analysis strategy in Figure 2.1.

It is interesting to compare this global ERP model with the structure of differences between the colour words derived on the basis of semantic similarity judgements (Figure 3.4a). Both configurations exhibit a polarity between BLACK and WHITE, and a somewhat circular structure having the order RED, YELLOW, GREEN, BLUE, PURPLE, RED. This relation seems very similar to that derived in the pilot experiment noted in the general introduction to this chapter. If the two structures are indeed related, then this relation would suggest that there is a semantic factor underlying differences between the colour name ERPs.

Structure derived from ERPs evoked by each of eight colour names

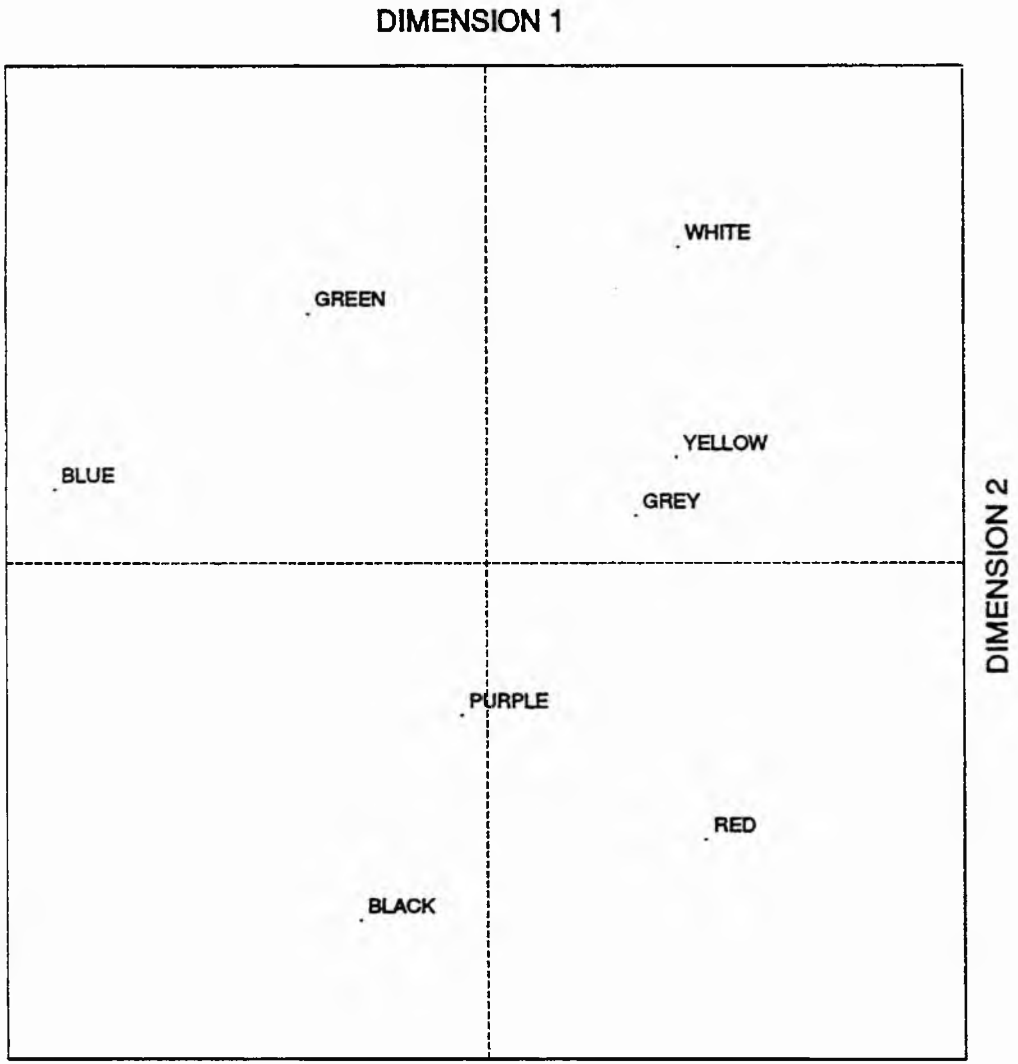


FIGURE 3.9

The human eye, however, is notoriously good at placing an inappropriate organization on visual input (Face of Elvis Seen On Mars!). This interpretability problem is particularly poignant here as a third dimension is completely hidden in the plane of the paper. For this reason I believe that these plots can be misleading, and so I intend to present no more ERP configurations: rather, the less misinterpretable variance-explained statistics will be presented as the sole indicants of relations between the linguistic and electrophysiological domains.

According to the procedure set out in Chapter 2 (see Figure 2.1), the global variance model was rotated by PROCRUSTES against each of the linguistic factor models. This produced a value of variance-explained for each model, which yields a quantitative insight into the goodness of fit between the ERP model and each linguistic factor. A large number of permutations of each factor model were then rotated against the ERP model in a randomization test designed to assess the approximate rarity of each variance-explained statistic.

Table 3.1 presents variance-explained statistics for the ERP global variance model, derived as previously described, for the entire epoch across the 6 subjects.

Only the semantic distance model fits well, explaining 0.53 of the ERP model variance. This goodness of fit approaches that between Cloze probability and N400 amplitude (Kutas and Hillyard 1984), as it is equivalent to a correlation of 0.73. The R^2 for this model is also statistically rare, being associated with a

Table 3.1

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-924 msec (6 subjects in experiment 1).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.049	.81
first letter freq.	.149	.47
last letter freq.	.053	.91
mean letter freq.	.046	.81
Word frequency models:		
K-F1	.022	.96
K-F2	.022	.95
K-F3	.034	.91
T-L	.028	.92
Word Length models:		
Number of letters	.101	.60
Number of syllables	.161	.32
Number of phonemes	.159	.32
Similarity Judgement models:		
phonological	.372	.19
orthographic	.419	.14
semantic	.526	.01

probability of 0.01. Thus a correspondence of this precision would be expected only once in a hundred permutations by chance. This probability is an order of magnitude less than that for any other model. For these data, it therefore may be the case that the more similar two words are judged to be in meaning, the more similar are the ERP signals evoked by exposure to them.

It has been remarked that ERP variance sensitive to the semantic aspects of verbal material may be distributed throughout the epoch, residing in no one ERP component (Molfese 1983; cf., Kutas and van Petten, in press). It does not seem very satisfactory, however, to be forced to conclude only that 'somewhere' in the epoch is variance related to a semantic factor. Additionally it may be that while the semantic factor is a major contributor to variance in the epoch, perhaps being the factor whose influence is most extended in time, it occludes other factors which have their effect on more temporally restricted portions of the waveforms. Accordingly, it was decided to examine more restricted time windows.

A natural point at which to segment the waveforms was just after the P200 peak. I thought it possible that an early window containing the N1-P200 phenomena might reveal aspects of the sensitivity of N1-P200 to semantic factors similar to those reported by Boddy (1981; 1986). The waveforms also appeared, by inspection, to reflect, prior to 250 msec, relatively simple modulations involving latency and amplitude differences which were similar at most electrodes. After this point, the waveforms appeared to be modulated in a more complex fashion.

These later modulations were thought either to be due to systematic differences in the spatio-temporal distribution of potential, or simply to noise.

Early ERP variance

An analysis based on a time window from stimulus onset to 250 msec was undertaken. Within-subject analyses of this shorter time window gave rise to higher variance-explained statistics than for analyses based on the whole epoch. The variance-explained statistics for each within-subject analysis averaged 0.94. The average R^2 for the analysis which collapsed across subjects was 0.51. Table 3.2 sets out the results of the analysis of the resulting configuration.

The variance in the global ERP model for this early window is best explained by models representing the low level features of the words, such as 'number of letters' (cf., Boddy 1981; 1986). 'Number of letters' explains 0.344 of the variance (corresponding to a correlation of 0.59) and is the only model which descends below an arbitrary rarity level of 0.05, having a probability of 0.04. It has a competitor, however: 'first letter frequency' also explains a good deal of variance and is associated with quite a small probability of having been obtained by chance. The values for the 'first letter frequency' model were 0.336 and 0.09 respectively.

Table 3.2

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-250 msec (6 subjects in experiment 1).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.074	.71
first letter freq.	.336	.09
last letter freq.	.104	.58
mean letter freq.	.099	.63
Word frequency models:		
K-F1	.037	.91
K-F2	.048	.82
K-F3	.016	.97
T-L	.079	.79
Word Length models:		
Number of letters	.344	.04
Number of syllables	.149	.47
Number of phonemes	.199	.18
Similarity Judgement models:		
phonological	.277	.57
orthographic	.205	.78
semantic	.069	.89

One way to conceptualise the fact that both models explain variance well is to note, referring back to Figure 3.7, that 'first letter frequency' occupies a region of the 'factor space' close to several word length-related models, including 'number of letters'. This letter frequency model is a covariate of word length. It could be that word length is the effective factor here and that the letter frequency model explains ERP variance in virtue of its shared variance with word length. Or it could be that the reverse is true. In this case, the word length model is a better predictor of ERP variance than is the letter frequency model, but the relative success in explaining ERP variance over the experiments reported below will be a better guide to which is more likely to have been the effective factor in these data. In either case, however, it seems that a relatively low-level feature of the words is the best predictor of variance before 250 msec.

Late ERP variance

Data drawn from a window from 250 to 900 msec were submitted to configurational analysis. As in the analysis of the early window, the within-subjects analyses explained more variance than the analysis based on the whole epoch. The R^2s averaged 0.82. The analysis undertaken to derive the global ERP variance model resulted in an average R^2 of 0.44. The resulting global ERP variance model was rotated by PROCRUSTES against each of the linguistic factor models, and the associated randomization tests performed. The results are summarised in Table 3.3.

Table 3.3

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 250-924 msec (6 subjects in experiment 1).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.061	.80
first letter freq.	.182	.18
last letter freq.	.140	.43
mean letter freq.	.171	.28
Word frequency models:		
K-F1	.010	.99
K-F2	.024	.97
K-F3	.011	.99
T-L	.054	.92
Word Length models:		
Number of letters	.122	.18
Number of syllables	.126	.45
Number of phonemes	.093	.57
Similarity Judgement models:		
phonological	.304	.35
orthographic	.462	.12
semantic	.570	.01

Only the semantic similarity model explains ERP variance well, and this it does to a level of statistical rarity of 0.01. The actual value of variance-explained is 0.57 (corresponding R = 0.75), a value slightly greater than that for the whole epoch solution. No other model is within an order of magnitude of the rarity of the correspondence for the semantic model.

Bearing in mind the caveat that no individual p-value be interpreted incautiously (see Chapter 2), this result suggests a relatively unequivocal interpretation. The suggestion is that there is a semantic factor underlying differences between ERPs evoked by exposure to the colour names, in so far that a semantic factor is captured by judgements of the semantic similarity of these words. The correspondence between these ERPs and the meaning of the evoking words seems confined to the later portion of the epoch: the analysis based on the pre-250 msec time window suggested a correspondence between ERP variance and relatively low level, letter-related aspects of the words. Hence, the sensitivity of variance in these ERPs appears to change at approximately 250 msec from low level, possibly physical, features to higher level semantic features.

Discussion of Experiment 1

The analyses carried out on the data recorded in experiment 1 suggest 3 candidate linguistic factors as important predictors of ERP variance. These are 'number of letters', 'first letter frequency' and 'semantic similarity'.

The analysis based on a window from stimulus onset to 250 msec suggested 2 candidate models for the underlying linguistic factor effective in this latency window. It would not be completely surprising if either word length or letter frequency were found to be factors underlying early variance. Several models of word recognition suppose that recognition proceeds by processing aspects of the lower level properties of a presented stimulus before representations related to their higher level properties are evoked. McClelland and Rumelhart (1981; 1985), for example, developed 2 related models of visual word recognition. In the first, the so-called 'interactive activation model' (McClelland and Rumelhart 1981), recognition occurs in a hierarchical architecture composed of neuron-like elements. Signals are first processed at a 'feature level', in which the elements of the stimuli which can be described in terms of local line segments at particular orientations are processed. The outputs of the neuron-like processors at this feature level innervate a higher 'letter level' at which particular constituent letters are identified. The output of the letter level is then fed to the highest 'word level', at which the word composed by the letters represented at the lower level is recognized. The system is interactive to the extent that representations at each level may not settle to a stationary state indicative of the features, letters and word best fitted to the input pattern, until signals from the higher levels have undergone feedback to the lower ones. The most likely word level solution may thus influence the most likely letter level solution. In a single processing episode, however, activity at the letter level will precede that at the level of word identification.

The later 'distributed model' (McClelland and Rumelhart 1985) supposes, as a general principle, that activation at all levels will be frequency sensitive. Sensitive, that is, to the number of previous processing episodes in which a particular feature, letter or word has been exposed. Frequently exposed letters will give rise to a larger activation across the neuron-like elements at the letter level (conceptualised as mean firing rate across the population of elements), and will do so more quickly, than low frequency letters (McClelland and Rumelhart 1985).

Both word length and letter frequency might therefore be thought possible candidates for the linguistic factors likely to be active soon after stimulus onset. It is slightly more difficult to rationalize, however, from the perspective above, why it should be 'first letter frequency' which explains early variance. Since the processing of all the constituent letters of the colour words is presumed to be frequency sensitive, it might be expected that 'mean letter frequency' would be the better predictor. The first letter alone is insufficient to distinguish between the words in the presented set.

The modulation of P200 amplitude by manipulations which involve changes to the physical aspect of stimuli are well known (Kutas and Hillyard 1983; Regan 1972). P200 is, for example, of greater amplitude for (repeated) physically larger word stimuli (Kutas and Hillyard 1983). A possibility, then, is that the word length effect in the early portion of the waveform is simply due to the fact that the long words are physically larger than the short words, and evoke a larger P200. If so, 'number of letters' is more likely to be the effective factor than is 'first letter frequency', since it captures this physical aspect of the stimuli (but see Chapter 6).

The sensitivity of ERP variance apparently changes from models signalling low level aspects of the words to a model which captures high level aspects of the words, at about 250 msec. The nature of this change, and its position in time after stimulus onset, suggested that it might be related to the processes by which semantic aspects of a word are generated from exposure to its graphemic representation. That is, it is possible that the variance modelled here may relate to the process of lexical identification or lexical access (see Monsell, Doyle and Haggard 1989).

Two of the candidate linguistic factors were associated with levels of statistical improbability which make it unlikely that the correspondences they signalled were due to chance. In particular the 'semantic similarity' model and the 'number of letters' model were both associated with p-values below the arbitrary 0.05 level. Notwithstanding the uncertainty regarding the effective linguistic factor in the early time window, the results of this experiment encouraged the view that the task of identifying linguistic factors underlying ERP variance might indeed be accomplished by the means developed in Chapter 2. A number of practical difficulties, however, were pointed up by conducting and analysing this experiment.

The first problem was that the way in which the data reduction and analysis procedures were organized on the mainframe computer resulted in lengthy and involved analyses. The procedure of analysing each subject's data separately, and abstracting information on average variance-explained and on the derived ERP configurations were particularly expensive in terms of interactive labour and computer printout. The practice of manually transcribing the within-subject variance models into a form suitable for the next stage of data reduction involved a great deal of time and made possible the intrusion of error. In some circumstances, due to many users taking up the processing resources of the mainframe computer, analyses such as those reported above could take more than a week of effort.

In order to accomplish the analyses more systematically, to reduce the amount of paper output, to reduce the possibilty of making errors in transcribing the data and to reduce the amount of time spent in interactive data processing (so that empirical data could be collected), I decided to 'automate' the data analysis procedures. This involved setting up batch processes which processed data from a selected time window without the need for any interactive work during the analyses themselves. The output could then be examined for processing errors and the results printed in the form of the PROCRUSTES and randomization test system logs. The control language for the portions of these batch processes which related to the important elements of the data analysis procedure are appended in Appendix 1.

This new organization of the procedures made possible more practicable analyses. A cost, however, associated with the more efficient organization of the analysis procedure was that the within-subject average variance-explained statistics, which varied little between analyses of comparably long latency windows, were not recorded. The analyses in which these statistics were embedded were not output, in order to save paper (on ecological grounds), but the system log of each analysis was carefully checked to see that these values were not eccentric. In the circumstance that an unexpectedly low or high value was present, the relevant analysis could be output. The variance-explained statistics associated with the MDS analysis which collapsed across subjects were also embedded in the analysis which, in the new organization, was not output. These values, however, could be reconstructed from the 'overall importance of each dimension' statistics which were always retained in the system logs.

A second concern, noticed while conducting these analyses, was that the 3-dimensional models were systematically better at explaining variance in the 3-dimensional ERP models than were the unidimensional models. This was not anticipated on the basis of the literature surrounding the employment of PROCRUSTES rotation (e.g. Alvey 1980). In the subsequent randomisation tests, however, every permutation of a 3-dimensional model will also explain more variance than will every permutation of a 1-dimensional model. The 3-dimensional judgement models must explain more variance in the ERP data in order to approach statistical rarity than must the unidimensional models. This aspect of the behaviour of the data analysis techniques, therefore, does not pose insuperable interpretative difficulties, but it does require to be noted.

A number of other practical problems centred on the passive exposure task employed in this experiment. With an ISI of 5 seconds, those who took part were subjected to a period of more than 45 minutes in which they had only to blink once every 5 seconds and try to attend to stimuli which did not engage them in any overt task. The paradigm was, in consequence, extremely tedious for the subjects.

The desirability of a task has been noted by several ERP researchers (e.g. Kutas and van Petten, in press; Neville 1980), for a number of reasons. First, without an overt task one cannot be sure that the subjects were processing the stimuli. Second, if the experimenter's interest is in the electrophysiological correlates of some cognitive process (see Chapter 1), then that process must be engaged by a task, and ERP measurements obtained in parallel with task performance (Kutas and van Petten, in press). In the present case, the process of interest is that of visual word recognition. The small possibility that variance related to the performance of a task might contaminate differences between ERPs evoked by each colour word seems less important than the certainty that subjects actually identify the presented words. Accordingly, the passive exposure paradigm is dispensed with in the following experiments and an appropriate task undertaken, in which, for the reasons noted above the ISI is reduced.

Experiment 2

Introduction

Despite the interesting results of experiment 1, it was decided to dispense with the passive exposure presentation paradigm. Bearing in mind, however, the desirability of emphasising split group techniques in exploratory studies (see Chapter 2; Muller, Otto and Benignus 1983) I wanted to try to retain as much commonality as possible with experiment 1. It was further decided, therefore, to employ a 'semantic decision', or 'category membership decision', paradigm in the experiments reported below. This required that subjects make a button press response with their preferred hand to orthographically similar non-colour word targets (see below) which were nested in the presentation sequence.

This task was intended to ensure that every word was processed at least to the level at which its identity became known. Processing to this level was thought necessary in order for the subject to respond to the out of category targets, or to withold response in the case of a colour name. Note that no response was required to the words of principal interest. This was intended to minimize the extent to which the colour word ERPs were contaminated by variance related to response execution processes (see Rugg 1987).

Method

Subjects: Six young adults were payed for participation. All subjects were right handed as determined by self report, were naive as to the purpose of the experiment, learnt English as their first language and possessed intact colour vision. None was being prescribed medication at the time of their participation and none had taken part in an evoked potential study before. Four subjects were male.

Design and task: Each subject undertook 4 sub-experiments which together comprised experiment 2. As in experiment 1, these were an ERP recording session and 3 similarity judgement tasks. The order of the tasks and of the ERP session were randomized, but half the subjects undertook a judgement task before the ERP experiment. All four tasks were undertaken within a single 2 hour experimental session.

ERP design and task: The ERP experiment involved each subject in a category membership decision task which entailed making a button press response to non-colour word targets which were placed at random intervals in the presentation sequence. The out of category targets were 100 words selected on the basis of their orthographic similarity to the set of colour words. Orthographic similarity was ensured by changing as few letters as possible of the colour words to make familiar non-colour associated words (e.g. ROD, BELLOW, GROAN, BACK etc.). The response was made with the preferred hand in all cases. The probability of the presentation of a target on any trial was 0.17, from which it follows that the colour words were 83% non-targets. No overt response was made to the colour words. In addition to the instructions relating to the category membership decision task, subjects were asked to try to withold blinking for the duration of a trial by allowing themselves to blink only while the fixation asterisk was displayed on the screen.

The decision was made to reduce the set of presented colour names to 10, by the exclusion of PINK. Thus, in factorial terms, the design was a word condition factor (with 10 levels) by an electrode channel factor (with 13 levels) with repeated measures.

ERP recording:

Electrodes: The electrode montage, the skin preparation procedure and the electrode impedance criteria in this experiment were identical to those in experiment 1.

Amplification: The amplifier settings and bandpass were identical to those of experiment 1.

Recording and digitisation: These factors, again, were identical to those which obtained in experiment 1. The response latencies, however, were recorded by another computer, in such a way that mean latency could be computed off-line, and behavioural errors associated with every trial (misses or false positives) investigated.

Stimulus presentation: The physical parameters relating to the presentation of the stimuli and the sequence of events which constituted one trial were the same as experiment 1 save that the ISI was reduced from 5 to 2 seconds. This reduction was in order to make the presentation less tedious for the subjects and to reduce the period of time necessary for conducting the ERP experiment. As noted above the colour word set was reduced to 10 by the exclusion of PINK. The presented set therefore comprised RED, ORANGE, YELLOW, GREEN, BLUE, PURPLE, BROWN, GREY, BLACK and WHITE. As before, there were 50 presentations of each colour name.

Averaging: Averaging was performed off-line for both the EEG data and for the reaction time (RT) data. In this experiment, not only artifact in the EEG and EOG data (such as blinks or movement related noise) but also behavioural errors were used to exclude trials from the averages. Thus a trial was only included in an average if no behavioural error occurred nor any extreme values in any of the electrode channels.

Procedure: After application and impedance testing of the electrodes, and, for half the subjects, performing one or more of the judgement tasks, subjects were seated in a soft upright chair in front of a TV monitor. The subject was given a hand held spheroidal 'bulb' on which was the thumb operated response button. The clavicular-vertebral reference electrode pair were balanced to remove heart artifact from the EEG trace. The task was explained, as above. A short practice sequence (20 trials of a sequence of colour names with targets which did not subsequently appear in the experiment proper) was administered to establish that subjects understood the task and had no problem seeing the words. When the subject was comfortable and ready, the experiment began. The presentation was composed of 6 blocks of 100 trials each, between each of which there was a short break. Subjects were debriefed during removal of the electrodes, which was always the last aspect of the experimental session (i.e. after any remaining judgement tasks).

Judgement task design: All three judgement tasks were identical to those performed in experiment 1.

Results and discussion

Behavioural data:

Reaction time and error rate: Mean reaction time to the out of category targets was 640.7 msec, with a standard deviation of 83 msec. The mean error of omission rate (misses) was 3%, and the mean false positive rate was 0.17%. These error rates were considered too low to analyse further.

Similarity judgement data: As described in experiment 1, judgement data from this experiment were put together with those from experiments 1 and 3, and employed to derive 3-dimensional MDS solutions. The coordinates of the 3 judgement models were employed as the semantic, orthographic and phonological models against which the ERP models derived from this experiment would be compared (see Appendix 2). These configurations were depicted in Figures 3.4 to 3.6. Figure 3.7 gives a useful guide to the interrelations between the linguistic models employed in this experiment.

Electrophysiological data

Figure 3.10a shows the grand average waveforms for each colour word across the 6 subjects who participated. Figure 3.10b shows the same waveforms, but only at the vertex electrode and at a greater scale. These waveforms are presented to show the general morphology of the traces only. There was a median of 48 trials constituting each subject's average ERPs. As far as possible each colour word ERP is plotted in the colour which each word denotes.

The morphology of these waveforms is more complex than those recorded in the passive exposure task of experiment 1. There are at least 7 deflections in the post-stimulus traces. There is a P1 deflection which peaks at about 50 msec, an N1 deflection peaking at about 140 msec, a P200 at about 210 msec, a negativity at about 300 msec (N280), a positivity at 400 msec (P400), a subsequent sharp negativity at 450 msec (N450) and a further positivity, which is maximal at about 600 msec (P600).

By comparison with the waveforms in experiment 1 (see Figures 3.4a and 3.4b), P200 is somewhat reduced, probably due to the overlapping presence of the subsequent biphasic negativity which peaks at ~280 msec (N280). It is not

Grand average ERPs evoked by each of 8 colour words in experiment 2.

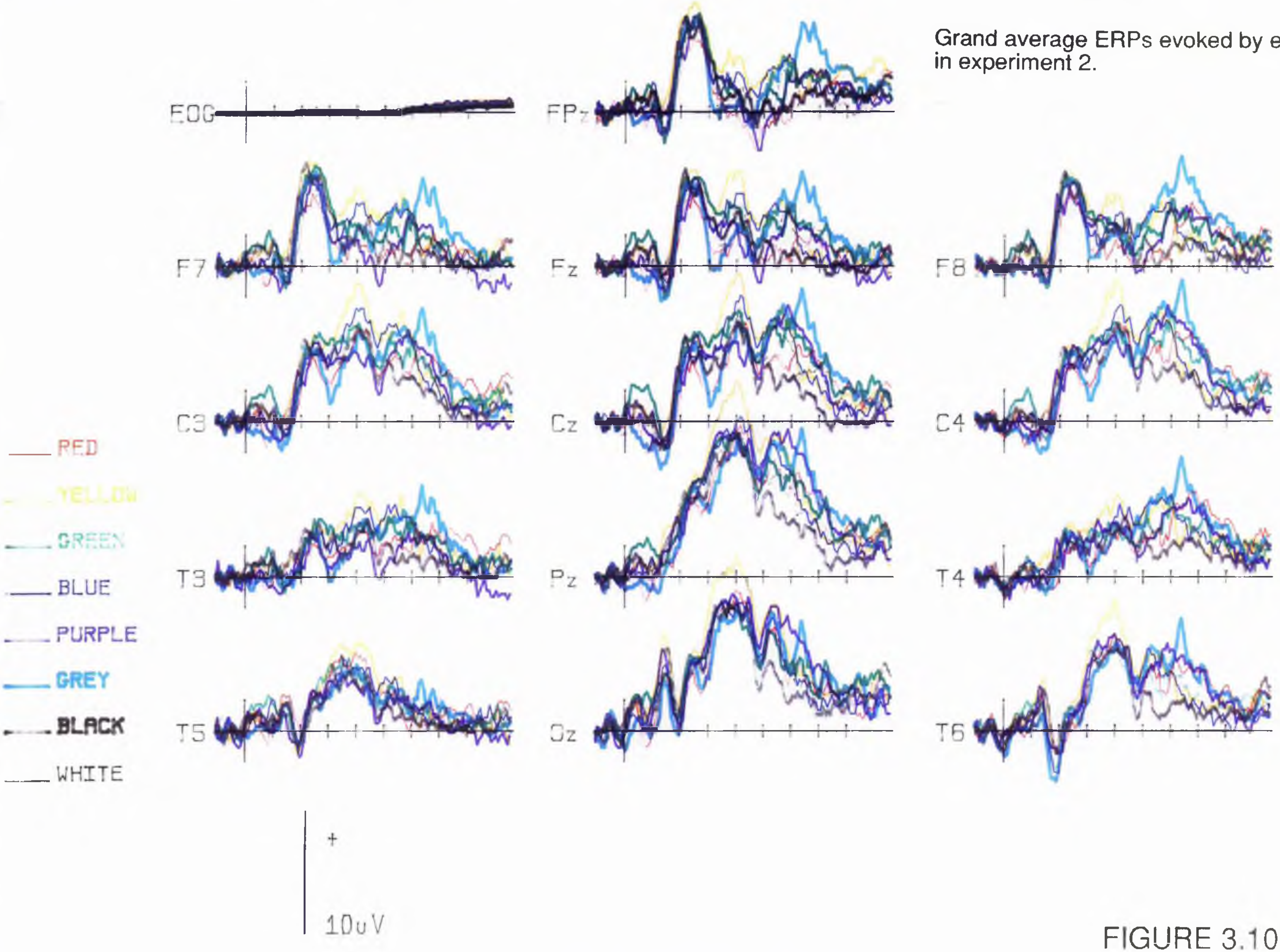
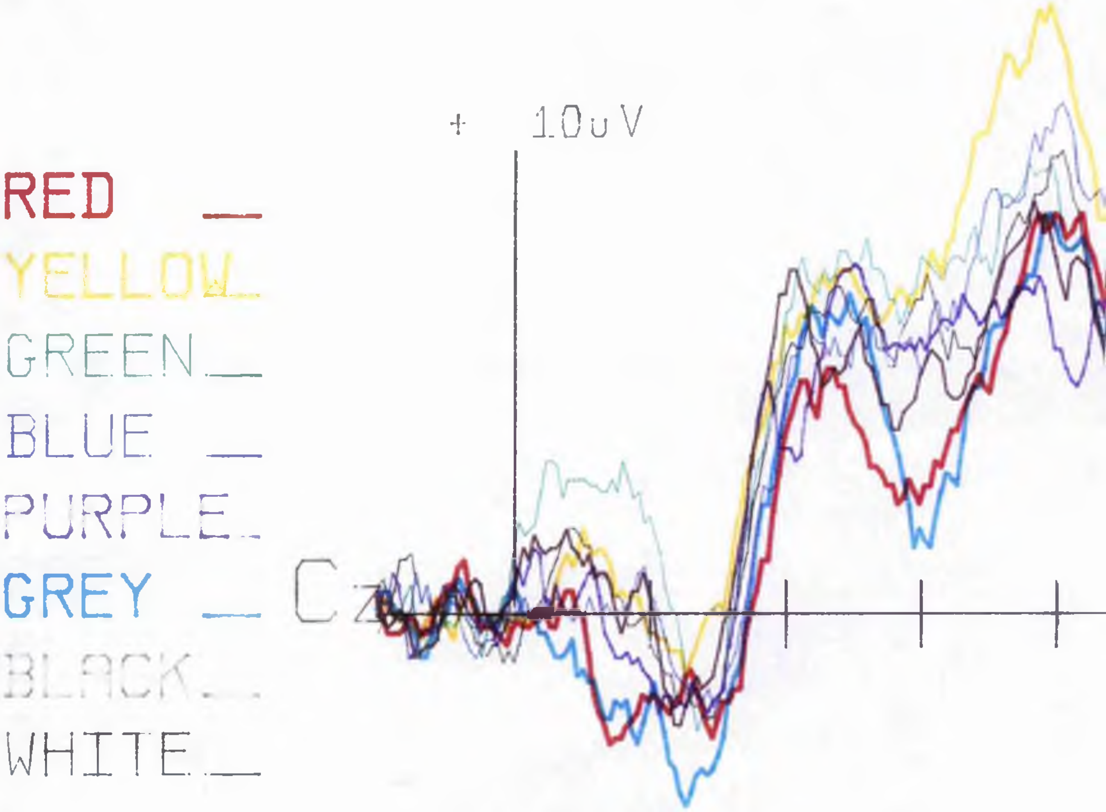


FIGURE 3.10a

Ticks are at 100 msec intervals

test using 100 iterations from the 10! possible permutations (Edgington 1980).



Ticks are at 100 msec intervals

Grand average vertex ERPs evoked by each of 8 colour words in experiment 2.

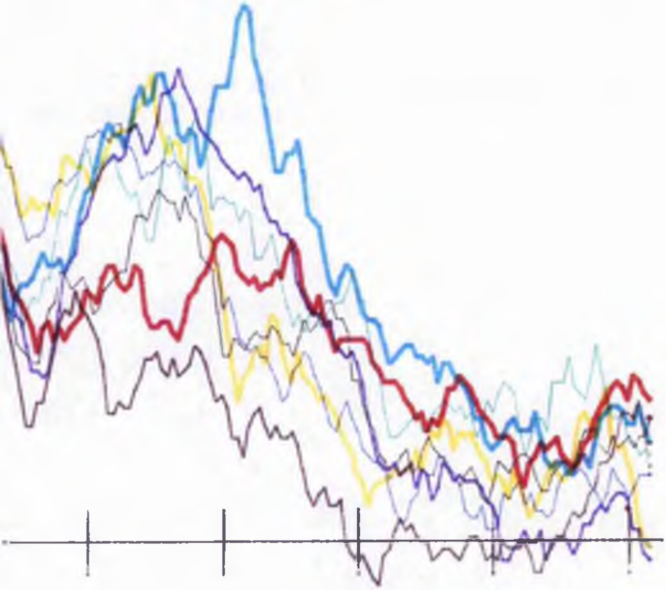


FIGURE 3.10b

obvious from inspection that the P400 which follows the N280 is a frontally distributed P3a (see Chapter 1). This component appears, rather, to have a parietal maximum, as does the later P600. It is possible that both P400 and P600 are part of the same positive going complex in which is embedded another process: the sharp negative going deflection at 450 msec (N450). In any case, the P600 is associated with the greatest amount of variance between conditions, and has the appearance of a visual P3b (see Chapter 1).

These data were submitted to configurational analysis. As before, analyses were based on responses to 8 colour names. An analysis based on the whole epoch was undertaken first. The variance-explained values of the within subjects analyses and of the across subjects analysis were comparable to those of the whole epoch analysis of experiment 1 (see the discussion of experiment 1). This analysis yielded high variance-explained statistics for all the word frequency models, which for Kucera-Francis 3 was 'significant' at $p = 0.01$ (see Table 3.4).

This result was very surprising for two reasons. First, some subjects (those who had performed all three judgement tasks before the ERP session) had been exposed to each of the words several hundred times in only two hours. It seemed to me most unlikely that information salient to word frequency would be preserved in the face of so many repetitions of the stimuli. Second, this result is completely different to that of the comparable analysis in experiment 1.

Table 3.4

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-924 msec (6 subjects in experiment 2).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.194	.13
first letter freq.	.071	.63
last letter freq.	.142	.27
mean letter freq.	.171	.20
Word frequency models:		
K-F1	.287	.03
K-F2	.270	.05
K-F3	.294	.01
T-L	.268	.05
Word Length models:		
Number of letters	.121	.36
Number of syllables	.027	.85
Number of phonemes	.018	.92
Similarity Judgement models:		
phonological	.259	.31
orthographic	.163	.55
semantic	.153	.59

My first intuition was that my laboriously developed analytic procedures were producing nonsense. By examining the waveforms, however, I noticed that the traces tended to change their relative positions just prior to the P600, at which point there was an almost monotonically ordered relation between low frequency words being positive, and high frequency words being negative (see Figure 3.10b). This suggested that the major influence on the waveforms at 600 msec might indeed be word frequency. The greater variance at this point in the waveforms might also be occluding variance related to the factors suggested by experiment 1 (see Chapters 5 and 8). Accordingly, more restricted time windows were submitted to configurational analysis.

Early ERP variance

Data from a time window from stimulus onset to 250 msec were analysed. The correspondences between the resulting global ERP variance model and the linguistic factor models are summarised in Table 3.5. Only one linguistic model explained variance to a degree which is statistically improbable by chance. 'Number of phonemes' explained 0.32 of the ERP model variance, and this variance-explained value was associated with a probability of 0.03. By consultation of Figure 3.7, it can be seen that this word length model occupies the region of the 'factor space' related to word length. Hence, as in experiment 1, a word length model is the best predictor of the structure of differences between the ERPs in the early post-stimulus epoch.

Table 3.5

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-250 msec (6 subjects in experiment 2).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.260	.10
first letter freq.	.119	.84
last letter freq.	.100	.60
mean letter freq.	.279	.07
Word frequency models:		
K-F1	.157	.47
K-F2	.132	.60
K-F3	.158	.47
T-L	.201	.25
Word Length models:		
Number of letters	.249	.12
Number of syllables	.118	.60
Number of phonemes	.320	.03
Similarity Judgement models:		
phonological	.162	.99
orthographic	.288	.57
semantic	.302	.15

Mid-latency ERP variance

I noted above that the traces appeared to reorganize their relative positions prior to the P600 deflection. Since it was suspected that differences between the ERPs in the latency region of the P600 were related to word frequency, I needed to decide on another latency point at which to segment the epoch. This latency should be so placed that the variance associated with the P600 is parcelled out. This was in order to investigate which linguistic factor, if any, was 'driving' variance in the latency region between 250 msec and the P600. To this end I chose to segment the waveforms at 400 msec. Accordingly a configurational analysis was applied to data from a window between 250 and 400 msec.

Table 3.6 summarizes the results of this analysis. The only model to correspond to the ERP model at a level of precision which was improbable by chance was 'semantic similarity'. This model explained 0.52 of the variance in this window, and is associated with a probability of 0.02. Only the word length model 'number of letters' is within an order of magnitude of the rarity of the result for the semantic model. Hence, as was found in experiment 1, 'semantic similarity' is the best predictor of variance in a window beginning at 250 msec. Similarly in accord with the results of experiment 1, there was apparently a change in the linguistic factor to which ERP variance was most sensitive, the change being from a low level factor to the semantic one at about 250 msec.

Table 3.6

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 250-400 msec (6 subjects in experiment 2).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.020	.99
first letter freq.	.106	.82
last letter freq.	.088	.56
mean letter freq.	.166	.35
Word frequency models:		
K-F1	.062	.82
K-F2	.022	.99
K-F3	.054	.85
T-L	.145	.49
Word Length models:		
Number of letters	.305	.07
Number of syllables	.216	.26
Number of phonemes	.132	.39
Similarity Judgement models:		
phonological	.175	.89
orthographic	.352	.28
semantic	.519	.02

Table 3.7

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 400-800 msec (6 subjects in experiment 2).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.072	.72
first letter freq.	.132	.54
last letter freq.	.176	.31
mean letter freq.	.076	.72
Word frequency models:		
K-F1	.321	.02
K-F2	.319	.04
K-F3	.351	.01
T-L	.287	.03
Word Length models:		
Number of letters	.081	.72
Number of syllables	.215	.14
Number of phonemes	.008	.97
Similarity Judgement models:		
phonological	.438	.17
orthographic	.427	.18
semantic	.090	.89

Late ERP variance

An analysis based on a time window extending 200 msec either side of the peak latency of the P600 (400 - 800 msec) was undertaken. This confirmed my suspicion, generated informally by inspection of the waveforms (see above), that it was this latency region which gave rise to the correspondence of the whole epoch ERP model to word frequency. Table 3.7 sets out the 'goodness of fit' of the ERP model for this window with each of the linguistic factors.

All four word frequency models explain variance well in the ERP model and are associated with low probabilities of their R^2's having been obtained by chance. No other model corresponds to the electrophysiological model at a probability level as low as those of the word frequency models.

Discussion of experiment 2

ERPs evoked by exposure to the colour names differed from each other in this experiment. The structure of these differences was found to change from orderings correlated with 3 linguistic factors as a function of time.

Prior to 250 msec, the ERPs were ordered according to what would be expected if word length were responsible for the differences between them. That is, the more similar two evoking words were in word length, the more similar were the ERP signals evoked by exposure to them.

Differences between the ERPs after this point, but before 400 msec, were ordered according to what would be expected if semantic similarity were affecting the waveforms. That is, the more similar two words were judged to be in meaning, the more similar were the ERP waveforms they evoked in this time window.

Variance centred on the P600 deflection was ordered according to word frequency. In other words, the more similar two words were in word frequency, the more similar were the ERPs they evoked in this time window.

For these data it seemed that, as in the passive exposure experiment (experiment 1), variance before 250 msecs was largely sensitive to low level aspects of the words, such as those captured by the word length variables, and that after that time point variance was sensitive to 'semantic' aspects. As 250 msecs is just short of the time that a competent reader of English invests in a content word in fluent reading (Carroll and Slowiaczec 1986; Just and Carpenter 1980), these results could be taken to imply that this technique might make aspects of the processes by which a word's meaning is recovered during reading accessible to electrophysiological inspection (but see Chapters 7 and 8).

Both experiments 1 and 2 suggest that a low level linguistic variable, most probably word length, is followed by a high level variable related to word meaning. This consistency can be seen to be unlikely to have come about by chance when put into the context of what *might* have happened. Any model, of 14 competing models, could have best explained variance in either the early or the later window of either experiment. Yet in both cases a word length model (of which there were 3) was followed by the semantic similarity model as the best variance predictors. Leave aside the fact that this result is obviously consistent with the expectations of most theoretical approaches to word recognition (see Chapter 7), as, for example, the opposite temporal relation between these two models would not be. The probability that this pattern of correspondence between the linguistic and ERP domains over the 2 experiments could be due only to chance is very small.

A correspondence between the ERP data and word frequency was present in the later 400-800 msec window. This correspondence seemed to be mediated by differences associated with the P600 deflection. There was neither a word frequency effect nor a P600 deflection in the waveforms evoked in experiment 1. This implies that the category decision task engages a late processor which is sensitive to word frequency. The fact that this word frequency effect is present even after multiple repetition suggested that any model of frequency effects which accounts for them in terms of local familiarity could not be true: even after all words are extremely familiar, some processor in the brain still respects a distinction between low and high frequency words. Possibly this reflects the intuition that repeating a low frequency word many times in a limited context does not abolish the knowledge that it is, in fact, a low frequency word (see Chapter 7).

Although not strictly related to the task of identifying the linguistic factors which underlie ERP variance, the component structure of these waveforms was interesting. Despite the fact that these ERPs were evoked by exposure to non-targets, the late positivity (P600) seemed much like a visual P3b (see Chapter 1). This raised the possibility that there were systematic differences in the subjective probability of the colour words, with low frequency words being felt to be less likely to occur in the presentation sequence than high frequency ones (see Chapter 7). An insight into this could be gained by recording subjects impressions of whether some words seemed to have been presented more than others.

The sharp N450 deflection seemed unlike other components previously reported as occurring at that latency, such as N400. N400 is a broad monophasic negativity (see Chapter 7), and not, like the N450, a sharp negative going deflection which appears to be over within 100 msec (see Figure 3.10b). The possibility that the N450 might be a stimulus offset transient was explored, though the presentation parameters are unchanged from experiment 1 and from most other studies from this laboratory (e.g. Rugg 1987; Rugg, in press), none of which show the presence of this deflection.

Another explanation might follow from the fact that the non-target words were so familiar to the subject. Rather than using a semantic criterion to restrain response to the colour words, subjects might be using criteria based on the shapes of the words, and therefore performing some sort of post-lexical letter level analysis (see Monsell 1985). This could have seemed more reliable than dependence on 'feeling' alone (perhaps being more cognitively penetrable (Pylyshyn 1984)). All subjects who were asked, during electrode removal, admitted that they were, indeed, using 'templates' of the familiar non-target words. Thus my suspicion is that the N450 deflection reflects this process. It would therefore be expected to diminish in circumstances which reduced the ability of subjects to use this strategy, such as those in which unfamiliar non-target category items are injected into the presentation sequence. But it might be noted that the predisposition of subjects to use a letter-level strategy, if that is what they are doing, could reduce the recoverability of variance sensitive to semantic factors in this paradigm.

Experiment 3

Introduction

At this point in the development of these studies it appeared to me that the analyses suggested that there were three linguistic factors underlying ERP variance in the experimental paradigms employed. These were the early sensitivity to word length, the change at ~250 msecs to sensitivity to a semantic factor and, in the category decision experiment, the late word frequency correspondence. In order to check that these linguistic factors were indeed the important ones, and to assess the degree to which the effects were replicable, it was decided to conduct another category decision experiment. This was designed to be in all respects similar to experiment 2.

Method

Subjects: Six young adults were payed for participation. Five subjects were female. All were naive as to the purpose of the experiment, all were right handed and possessed normal colour vision, as determined by self report, and all learnt English as their first language. Three subjects had prior experience of an ERP experiment. None were taking prescribed medication at the time of the experiment.

This experiment is in all other methodological aspects identical to experiment 2.

Results and discussion

Behavioural data:

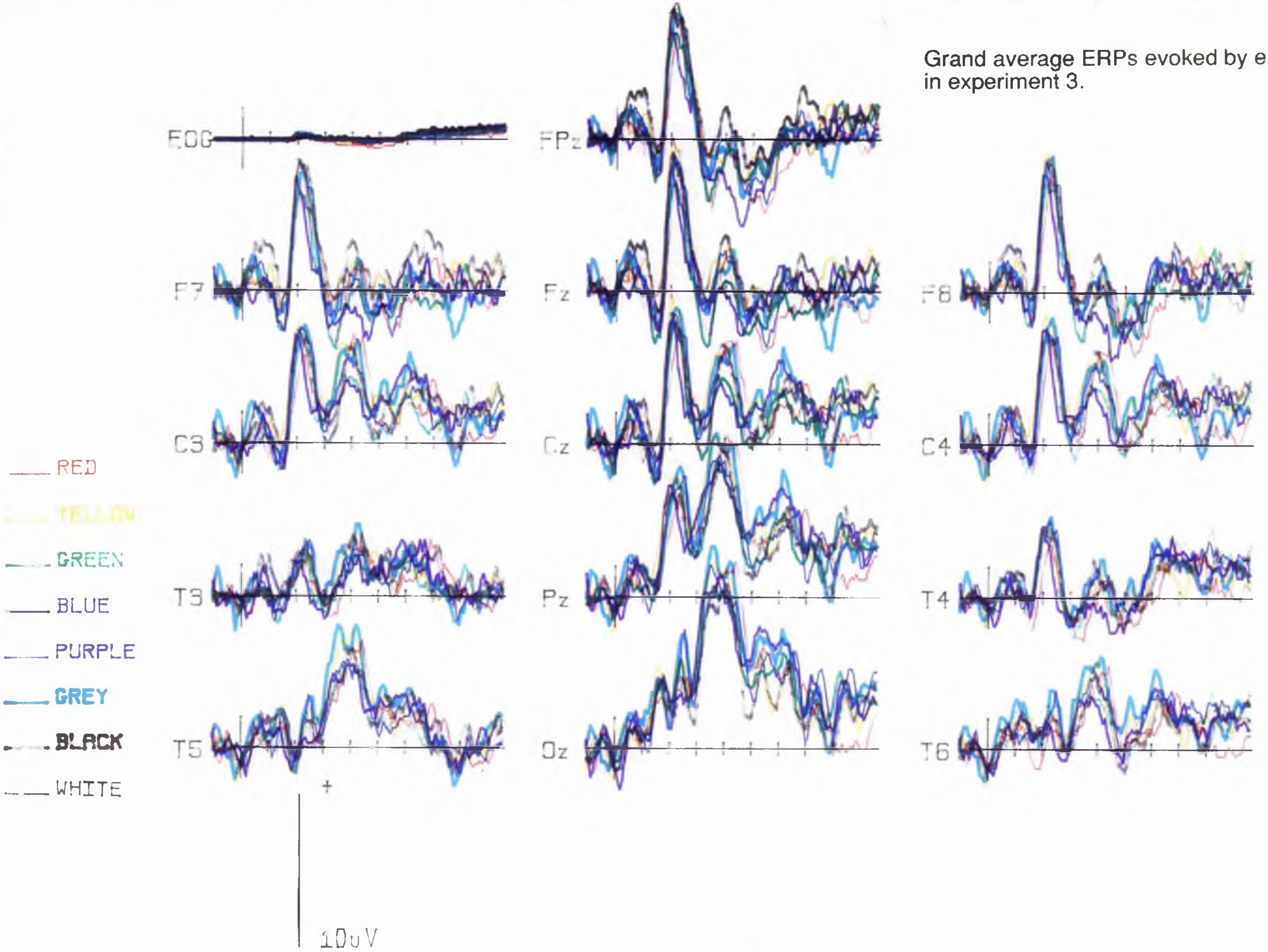
Reaction time and error rate: Mean reaction time to the out of category targets was 593 msec. The standard deviation was 67 msec. The error rate for misses was 3.3%, and the error rate for false alarms was 0.6%. These rates were considered to small to analyse further.

Similarity judgement data: As detailed in experiment 1, 3-dimensional MDS solutions were computed for each of the 3 judgement data sets. The resulting configurations were depicted in Figures 3.4 to 3.6. If it is considered useful, Figure 3.7 may be referred to as a guide to the inter-relations between the linguistic factor models employed in this experiment.

Electrophysiological data

Figures 3.11a and 3.11b present the grand average ERPs which were evoked by each colour name in this experiment. The median number of trials averaged to produce each subject's ERPs was 49. The general morphology of these waveforms was very similar to those recorded in experiment 2, indeed there were no important differences between the sequence of deflections recorded in the two subject groups.

Grand average ERPs evoked by each of 8 colour words in experiment 3.



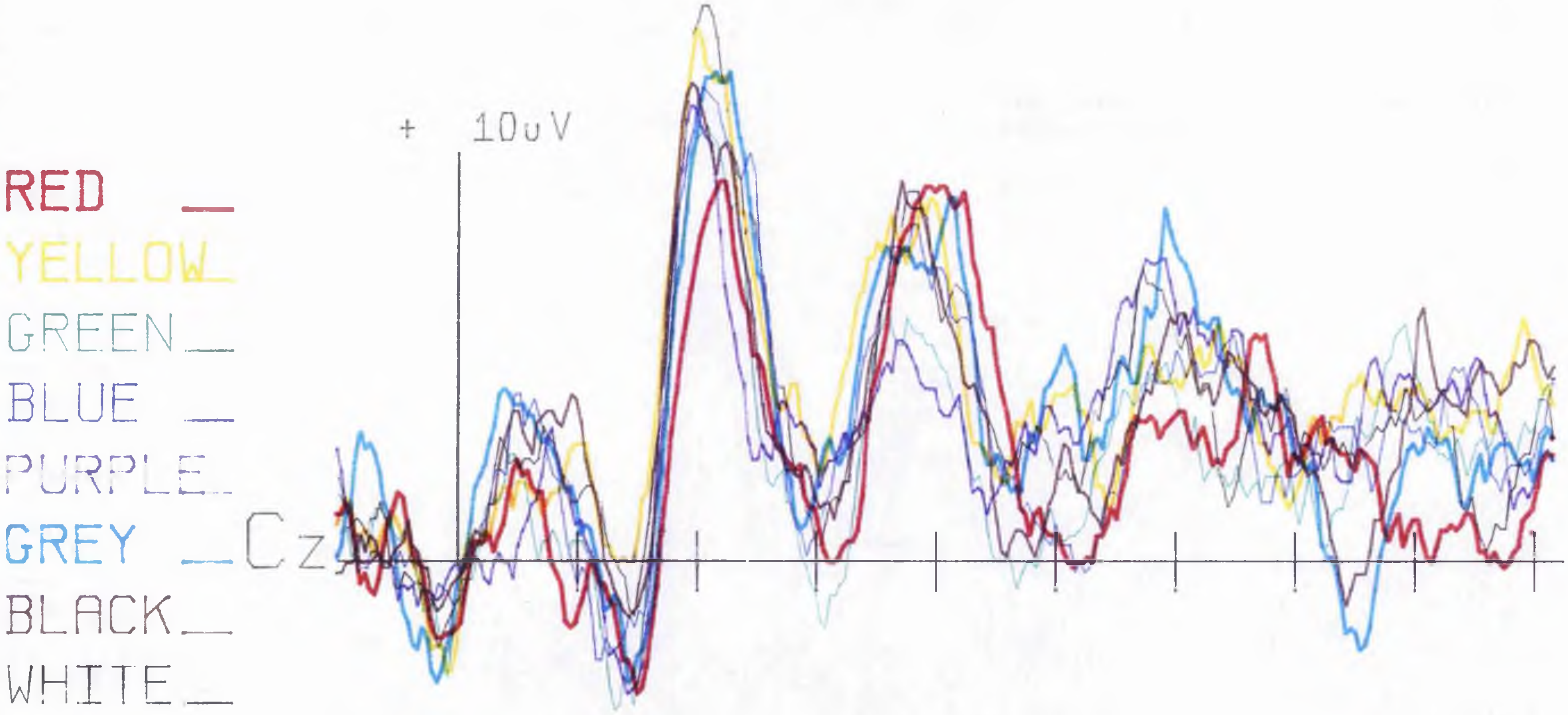
Ticks are at 100 msec intervals

FIGURE 3.11a

a balanced sternovertebral pair. Words were presented in pseudorandom order, each occurring 50 times.

In addition to the ERP recording session, subjects also participated in 3 similarity judgement tasks. These were designed to allow computation, by psychological scaling (Shepard 1980), of models for the words orthographic, phonological and semantic differences. The order of these tasks and of the ERP session was randomized.

Grand average vertex ERPs evoked by each of 8 colour words in experiment 3.



Ticks are at 100 msec intervals

FIGURE 3.11b

Configurational analysis was applied first to ERP data from the whole epoch. Table 3.8 summarizes the results.

There is a tendency for ERP variance over the whole epoch to be sensitive to word frequency. This is reflected in the fact that 'Kucera-Francis 2' is the model which corresponds to the ERP model with the least probability that the relation is due to chance, and in the fact that the word frequency models in general are associated with lower probabilities than the other models. The probability associated with 'Kucera-Francis 2', however, is only 0.09. While this aspect of the results is qualitatively similar to that of the whole epoch analysis of experiment 2, the relation between ERP variance and the linguistic factor models is not as clear cut.

Configurational analyses based on time windows exactly analogous to those of experiment 2 were undertaken.

Early ERP variance

Table 3.9 summarizes the results of an analysis on data drawn from a time window from stimulus onset to 250 msec. The orthographic similarity model is well fitted to the ERP model (0.549) and associated with quite a low probability (0.07).

Table 3.8

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-924 msec (6 subjects in experiment 3).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.201	.27
first letter freq.	.204	.25
last letter freq.	.149	.50
mean letter freq.	.121	.55
Word frequency models:		
K-F1	.189	.28
K-F2	.259	.09
K-F3	.223	.12
T-L	.177	.29
Word Length models:		
Number of letters	.105	.63
Number of syllables	.170	.40
Number of phonemes	.054	.78
Similarity Judgement models:		
phonological	.308	.48
orthographic	.177	.87
semantic	.079	.88

Table 3.9

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-250 msec (6 subjects in experiment 3).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.000	1.00
first letter freq.	.169	.29
last letter freq.	.044	.87
mean letter freq.	.121	.54
Word frequency models:		
K-F1	.099	.65
K-F2	.035	.85
K-F3	.080	.70
T-L	.091	.71
Word Length models:		
Number of letters	.147	.45
Number of syllables	.197	.26
Number of phonemes	.122	.47
Similarity Judgement models:		
phonological	.363	.31
orthographic	.549	.07
semantic	.128	.71

In the space defined by Figure 3.7, the ERP model will be close to this 'orthographic similarity' model. As discussed in experiment 1, the 'orthographic similarity' model is in a region of the space depicted in Figure 3.7 which seems related to word length. The identification of this region with word length was made on the dual grounds that both 'number of letters' and 'number of phonemes' are present, and that part of the organization of the 'orthographic similarity' model seemed to reflect the influence of word length. It is not inconsistent, then, to suppose that ERP variance in this early window is, again, best fitted by a model related to word length. Nonetheless, inference from these data would have been less equivocal had one of the nominal word length models explained most variance.

Mid-latency ERP variance

Data from a time window extending from 250 msec to 400 msec were analysed by the configurational analysis procedure. The results are summarised in table 3.10.

All the models explain variance in the ERP model for this time window at a level which is quite likely to have come about by chance. There is no definite evidence for the action of any linguistic factor in this analysis. The only point to be made is that the 'semantic similarity' model is the one best fitted to the ERP variance (but not the least likely to be related to the ERP model by chance). This is what would be expected on the basis of the results of experiment 2, but this correspondence could be due to the action of chance.

Table 3.10

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 250-400 msec (6 subjects in experiment 3).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.076	.77
first letter freq.	.145	.36
last letter freq.	.084	.70
mean letter freq.	.094	.67
Word frequency models:		
K-F1	.152	.40
K-F2	.238	.14
K-F3	.198	.23
T-L	.200	.21
Word Length models:		
Number of letters	.099	.68
Number of syllables	.074	.70
Number of phonemes	.107	.51
Similarity Judgement models:		
phonological	.205	.81
orthographic	.164	.56
semantic	.319	.43

Late ERP variance

Table 3.11 presents the results of a configurational analysis performed on ERP data from a window between 400 and 800 msec. In common with every analysis of this experiment, all the variance-explained statistics are associated with a greater probability that they might have come about by chance. The word frequency models, however, are associated with lower probabilities than are any other models.

Referring once more to the 'factor space' shown in Figure 3.7, the point representative of the relation of the ERP model to the linguistic factor models will be at some position in the space. Noting that 'phonological similarity' and 'mean digram frequency' also explain some variance, the ERP model is likely to be at a point slightly to the left of the centre, and toward the lower region of the plot. This position is at least in the right region of the factor space, on the expectation that word frequency is the major factor at this latency in this task. The inference, however, that word frequency is effective here is scarcely compelling.

Table 3.11

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 400-800 msec (6 subjects in experiment 3).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.149	.23
first letter freq.	.078	.60
last letter freq.	.104	.49
mean letter freq.	.065	.68
Word frequency models:		
K-F1	.172	.19
K-F2	.129	.31
K-F3	.161	.21
T-L	.178	.14
Word Length models:		
Number of letters	.036	.82
Number of syllables	.135	.30
Number of phonemes	.036	.79
Similarity Judgement models:		
phonological	.311	.23
orthographic	.121	.77
semantic	.078	.91

Discussion of experiment 3

The results of this experiment, over all the time windows which were analysed, were characterised by the linguistic factor models showing a diminished facility differentially to explain ERP variance. In the hope of being able to account for this feature of the data, I decided to investigate what relation, if any, these results bore to those of experiment 2.

One way to approach this problem was to regress the variance-explained statistics across all windows for these data, on the corresponding values in experiment 2. If, despite the above problem, this experiment 'tells the same story' as experiment 2, then the pattern of variance-explained values across the time windows in experiment 2 would be expected to be a good predictor of the pattern in the present analyses. The input to the regression could be thought of as the list of R^2 values from tables 3.8 to 3.11 'stacked' into a single column vector, and then being regressed on a similar 'stack' derived from tables 3.4 to 3.7.

The result of this regression analysis, expressed as an F-ratio, was that the regression coefficient was significant ($F1,68 = 5.59$, $p = 0.02$). This result suggested that the *pattern* of variance sensitivity was similar in both experiments. This is to say that the word length models tend to be better at explaining early variance than variance in other windows, that the word frequency effects tend to be later, and so on. Some factor, though, which was more apparent in the data of experiment 3, acted to suppress the 'explainability' of the ERP variance.

One explanation, the one I prefer, is that this factor was noise: the ERP models in experiment 3 were more noisy, leading to the lower values of the variance-explained statistics (see below).

Immediately following the ERP recording session, before either a subsequent judgement task or removal of the electrodes in appropriate cases, subjects were asked if any words had appeared more often in the presentation sequence. If so, the experimenter wrote down those that appeared more often, according to the particular subject. Additionally, the subjects were asked to recall as many of the colour words as possible, while they were written down by the experimenter.

The possibility that the P600 word frequency effect of experiment 2 might be related to systematic differences of subjective probability between high and low frequency words was discussed in experiment 2. The relation between frequency and words appearing to occur more often was investigated using the reports gathered as above. Subjects did indeed, in 4 of the 6 cases, think that some words had been presented more often. But the words named differed between subjects, and bore no clear relation to frequency.

Only 2 of the 6 subjects were able to recall all 10 of the colour names. The recalled colour names differed between subjects.

General Discussion

At this point in the development of these studies I considered that there were 3 possible interpretations of the success with which the modelling technique was providing insight into the question of which linguistic factors (if any) underlie ERP variance. These interpretations varied from the very pessimistic to the cautiously optimistic.

First, and most pessimistically, it was possible that the results of all three experiments (not counting the pilot study) might have been illusory, and there are simply no effects present (besides chance). This interpretation faces a number of difficulties. It is difficult to account, by invoking only the presence of random interactions, for the clear consistencies between the first 2 experiments, and for the fact that a very similar pattern of variance explanation was found in experiment 3. If chance alone were operating, it would have been very unlucky indeed to turn up spurious results of the character reported above. Without attempting to calculate the odds against these consistencies, it can be seen that this interpretation just is not probable.

The second possibility was that there were systematic relations between the ERPs and the linguistic factors which the colour words embody, but that the data analysis procedure developed for these experiments might be inadequate to the task demanded of it. Perhaps it was too unstable under the effects of noise. This might be the explanation for the absence, for the most part, of clear effects in experiment 3. The data analysis procedures, however, have been responsible for the detection of any and all of the effects described above. No analysis which proceeds by fitting empirical variance could be expected to identify important factors when too great a part of the variance is due to noise.

A third possibility was that there were effects present, the methodology was broadly sound, but that the data collected in experiment 3 were too noisy to produce the clear results seen in experiments 1 and 2.

On this latter interpretation, the analyses have detected in ERPs evoked by individual colour words the effects of 3 linguistic factors; word length, word frequency and semantic similarity. The same factors were acting in experiment 3 but were embedded in greater noise. In support of this interpretation, one could note the following pattern of consistency. In the early time window of every experiment, arguably, a word length-related factor explains ERP variance best. This is followed in every experiment by a window in which the semantic similarity model explained variance best. In the late window of the 2 category decision experiments (experiments 2 and 3), word frequency is the factor least likely to be related to the ERP variance by chance. Hence, in each time window of experiment 3 the 'right' models may have been best fitted to the ERP variance. But it would be advantageous, if this interpretation were the correct one, to generate some insight into the problems of noise associated with experiment 3.

The effects of noise in the ERP data will be to increase the diameter of the 'error spheres' (see Chapter 2) around each point in the ERP models. Thus the computed points will be poor estimates of the 'real', noise-free, points. The consequence of this will be that a linguistic model, which represents a linguistic factor which is effective in the time window to which the ERP model relates, will be able to explain less variance in the ERP model. Hence noise will tend to suppress the variance-explained statistics associated with the effective model, increase the associated probability that its precision of fit could have come about by chance, and so reduce the likelihood that the model will 'stick out' from the other competing models. In this respect, the respect of the detectability of systematic vartiance in noise, the data analysis technique is similar to any other inferential statistical approach.

The presence of noise perturbation in the data complicates the inferences which can be made from these analyses. In the context of the exploratory nature of Part One, I considered that a greater reliance on across-sample consistencies was an appropriate response to this problem. The expectations generated by an open-minded interpretation of individual p-values in the explorations in Part One can be tested rigorously in Part Two.

Part of the noise problem, I believe, lay in the employment of the non-cephalic clavicular-vertebral reference electrodes. This technique was used because the non-cephalic reference is said to be influenced less by intracerebral activity than are comparable reference sites, such as linked ears or the tip of the nose (Stephenson and Gibbs 1951). The basis of the balanced non-cephalic technique is that anterior and posterior points at the base of the neck are influenced by approximately equal but opposite cardiac potentials. The residual cardiac potentials may be removed by adjusting a 20,000 Ohm potentiometer until the electrocardiogram (ECG) disappears from the EEG trace (Stephenson and Gibbs 1951). The extent to which this removal of heart artifact is possible in practice is dependent on the orientation of the heart in the subject's chest and on finding the right balance with the potentiometer.

A complicating factor here was that the potentiometer was next to the subject, and therefore at some remove from the EEG machine, which was in the control room. Adjustment to remove heart artifact therefore required the experimenter to examine the EEG trace, go through into the experimental room, adjust the potentiometer, go through to the control room, observe the effect of this adjustment on the EEG trace, decide what further adjustment to make, go through to the experimental room, and so on. In practice there must be a limit to the length of time for which a subject can be kept waiting (in an already long experimental session) while the non-cephalic reference is adjusted in this fashion. In several cases, inevitably, there was residual heart artifact in the EEG trials which constituted the ERPs.

In order to remove this possible source of noise, the experiments which follow employ linked mastoid reference electrodes. This removed much of the difficulty associated with the non-cephalic reference at the cost of making the ERP waveforms slightly less sensitive to intracerebral activity, particularly that generated in the temporal lobes (Stephenson and Gibbs 1951).

The analyses undertaken on the 3 experiments above suggest, as noted, that there are 3 linguistic factors which are identifiable as being effective predictors of variance between ERPs evoked by individual colour words. For any small set of words, however, there are adventitious relations between the linguistic models which describe the ways in which they differ. Every small set of words will have a different 'factor space', like that for the colour names which is shown in Figure 3.7. (Remember that the set of words has to be small in order to allow experiments of practicable length: see Chapter 2).

In order to resolve the issues raised by the colour name experiments, and to determine whether the 3 identified linguistic factors are more generally effective at predicting ERP variance, it is necessary to examine the relations between ERPs evoked by a different set of words and the linguistic factors which they embody.

Chapter Four

An electrophysiological and behavioural investigation of the furniture terms

Introduction

This chapter presents empirical results of the application of the quantitative modelling technique developed in Chapter 2. The technique is applied to individual words selected from the category of furniture terms, which were exposed in a category membership decision paradigm identical to that used in experiments 2 and 3.

This category was selected somewhat arbitrarily, but part of the reason for the choice was that two of the linguistic factors identified in Chapter 3 as predictors of ERP variance dissociate these terms particularly well. The furniture terms differ widely in word length (e.g. BED to WARDROBE), and in word frequency (e.g. SIDEBOARD to TABLE, 1 to 198 occurrences, respectively, in Kucera and Francis (1967)). This variance was considered helpful to the task of identifying whether these factors were important to ERP differences, for a reason which can be made obvious by considering the absence of such variance: If word length were an important underlying factor in ERP differences, and all the words were of the same length, the influence of the factor would not be seen (see Chapter 8).

In contrast to the colour names, the psychological properties of furniture terms have not been extensively investigated. The category probably has fewer members than the colour name category, evidenced by the fact that the following 10 furniture terms largely exhaust the members of this category in relatively common use: CHAIR, TABLE, STOOL, BED, WARDROBE, SIDEBOARD, DESK, COUCH, SOFA, CUPBOARD (cf., Chaise-longue, escritoire, secretaire, bureau, vitrine etc.).

This set of 10 furniture terms was the one employed for the following experiments. I could see no principled reason for the exclusion of any members of this set from the analyses (cf., Chapter 3), and so all the analyses were conducted on the full set of 10 words.

In order to preserve continuity and comparability with the colour name experiments, essentially the same set of linguistic factors were employed. 'Thorndike-Lorge', however, which was one of the word frequency models, appeared to be redundant in the foregoing experiments, so the ERP models were not rotated against this model.

The numerical models designed to capture the ways these words differ, according to each of the selected linguistic factors, were constructed as before. The letter frequency models were derived by consultation of the Baddeley, Conrad and Thompson (1960) normative letter frequency data. The word frequency models were derived from the Kucera and Francis (1967) study. The word length models were generated by counting the letters, syllables and phonemes of the words (see Chapter 3). The phonological, orthographic and semantic similarity models were derived by MDS from judgement data recorded from each subject during the experimental session. The numerical values of all these models, which were, of course, quite different to those for the colour names, are attached in Appendix 3.

In accord with the split-sample techniques recommended by Muller, Otto and Benignus (1983), the subjects who took part in the furniture term sessions were divided into two groups. Subjects were assigned to the groups on the basis of whether they were numbered even or odd in the list of those who participated. Data from the two groups of subjects were analysed seperately. The two data sets are designated experiment 4 and experiment 5, both for clarity, and to follow consistently from the experiments in Chapter 3.

Experiment 4

Method

Subjects: The data set for this experiment was derived from seven young adults who were payed for participation. These seven subjects were a subset of the even-numbered subjects who took part. Two even-numbered subjects were rejected from the data set. One was rejected on the ground that it was discovered after the experiment that she had not learnt English as her first language. The other was rejected because it was subsequently learnt that she was being prescribed psychotropic medication. The following features are drawn from the questionnaires filled in by each the 7 participants who were included. Five subjects were female. All were right handed, as determined by self report. All subjects learnt English as their first language. All were naive as to the purpose of the experiment. None were being prescribed medication at the time of the experiment. Four had taken part in an evoked potential study before.

Design and Task: Each subject took part in 4 sub-experiments which together comprised experiment 4. As before, these sub-experiments were an ERP recording session and 3 similarity judgement tasks. The order of these tasks was randomized, but 3 of the subjects took part in at least 1 judgement task before the ERP experiment. All four tasks were accomplished in a single 2 hour experimental session.

ERP design and task: The ERP experiment involved each subject in a category membership decision task. This entailed making a button press response to out of category targets which were placed at random intervals in the presentation sequence. The out of category targets were 100 words selected on the basis of their orthographic similarity to the set of furniture words. Orthographic similarity was ensured by changing as few as possible letters of the furniture words to make familiar words (e.g. SODA, COACH, BAD, CHOIR etc.). The response was made with the preferred hand in all cases. The probability of the presentation of a target on any trial was 0.17. No overt response was made to the furniture words. Subjects were asked to try to blink only while the fixation asterisk was displayed on the screen, which was between trials. In factorial terms, the design was a word condition factor (with 10 levels) by an electrode channel factor (with 11 levels, see below) with repeated measures.

ERP recording:

Logistical changes in the laboratory made necessary a number of changes to the ERP recording criteria from those which obtained in the colour name experiments.

Electrodes: Burden Ag\AgCl electrodes were fixed with collodion at 11 10-20 system sites; FPz, Fz, Cz, Pz, Oz, F7, F8, C3, C4, P3 and P4, and were referred to linked mastoid electrodes (Cooper, Osselton and Shaw 1974). The decision to change the reference from a balanced clavicular-vertebral pair to linked mastoids was discussed at the end of Chapter 3. The electrodes for the mastoid reference are sited on the mastoid processes behind each ear. Scalp electrodes placed at temporal sites were thought to be too close to the reference electrodes to pick up useful activity. Hence electrodes previously placed at T5 and T6 were moved to P3 and P4, respectively. The amplifiers used in the present (and all following) experiments had only 11 channels, so a reduction of 2 channels was accomplished by discarding electrodes previously placed at T3 and T4.

The placements of the ground and EOG electrodes were unchanged, as were the skin preparation and electrode impedance criteria.

Amplification: All signals were amplified by DIGITIMER amplifiers (50 uV/V for the EEG channels, and 200 uV/V for the EOG electrodes). The amplifiers were again set to a high pass of 32 Hz and a time constant of 5 seconds.

Recording and digitization: The signals were digitized at a rate of 4 msec per point (250 Hz) by a CED 1401 laboratory interface at 12 bit resolution. The sensitivity of the digitisation was equivalent to 0.125 uV per digitisation increment for the EEG channels, and to 0.5 uV per digitisation increment for the EOG channel.

The recording epoch again began 100 msec before stimulus onset and continued for 924 msec thereafter.

Stimulus presentation: All physical aspects of the presentation were as detailed in Chapter 3. The ISI was 2 seconds, as for experiments 2 and 3. Ten furniture names: COUCH, TABLE, SOFA, CUPBOARD, SIDEBOARD, CHAIR, STOOL, WARDROBE, DESK and BED were presented 50 times in pseudorandom order.

ERP averaging and procedure: All aspects of the experimental procedure and the averaging process were identical to those for experiments 2 and 3, which were detailed in the method section of experiment 2.

Judgement task design and procedure: Subjects performed the 3 judgement tasks exactly as before, save that their judgements were of the relations between the 10 furniture names.

Results and Discussion

Behavioural data

Reaction time and error rate: Mean reaction time to the targets was 719 msec with a standard deviation of 94 msec. The mean error of omission rate was 2.25%, and false positive rate was 0.45%. These error rates were not further analysed.

Similarity judgement data: The semantic similarity judgement data collected in this experiment were put together with those of experiment 5, for exactly the same reasons as were detailed in Chapter 3. The combined data set (N = 15) was submitted to non-metric multidimensional scaling by INDSCAL (Carroll and Chang 1970).

The decision about how many dimensions to use to express the interrelations between the stimulus objects was made in the following way. As in the colour name studies, the decision was not informed by the results of a 'scree test', which showed no 'elbow' signalling a dimensionality above which added dimensions gave diminishing returns in terms of variance-explained. Nor was the decision helped by the number of reliably estimable dimensions (cf., Shepard 1972), which was again high. The decision was made on the basis of the interpretability of the recovered configurations. Solutions of dimensionality higher than 3 were characterised by the higher dimensions being very difficult to ascribe to any interpretable aspect of the words. As before then, the MDS solution for the semantic similarity judgements was computed in 3 dimensions.

The computed configuration in 3 dimensions explained an average of 0.807 the variance in the input judgement matrices. All subjects' judgement data were well fitted by the MDS model.

Figures 4.1a to 4.1c present the configuration of points which represent the interrelation of the furniture terms derived from judgements of their semantic similarity. Examining Figure 4.1a first, there seem to be 3 clusters of terms which relate to sub-categories with different uses. STOOL, CHAIR, COUCH and SOFA are grouped closely and appear to reflect a sub-category related to furniture items used for sitting on. The proximity of WARDROBE, CUPBOARD and SIDEBOARD reflects the fact that these terms refer to articles of furniture used for putting things in. DESK and TABLE are both used for putting things on, and are placed together about midway between the 'put things in' group and the 'sit on' group. BED is quite close to the 'sit on' group. Dimension 1 may relate to a 'put in' to 'sit on' dimension.

Structure derived from semantic similarity judgements of the furniture terms

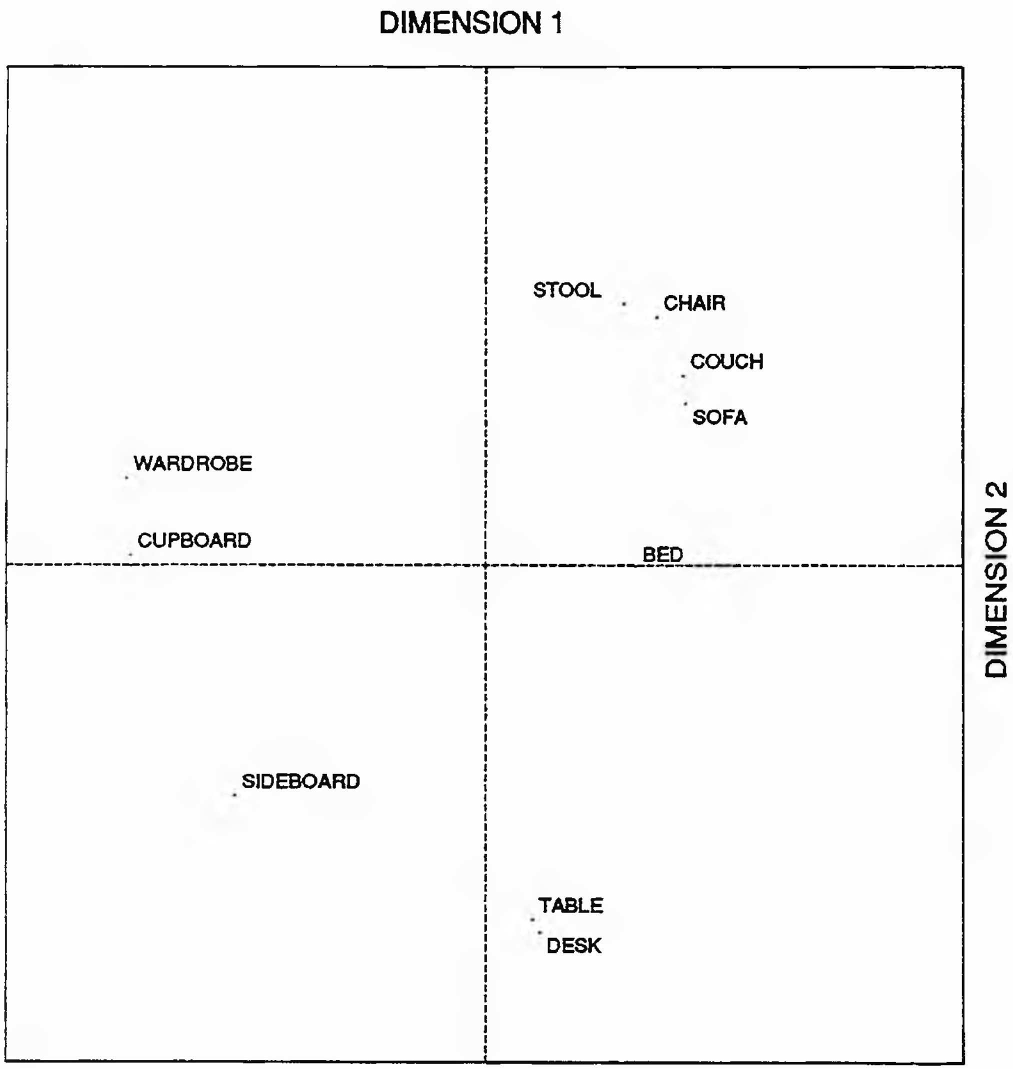


FIGURE 4.1a

Structure derived from semantic similarity judgements of the furniture terms

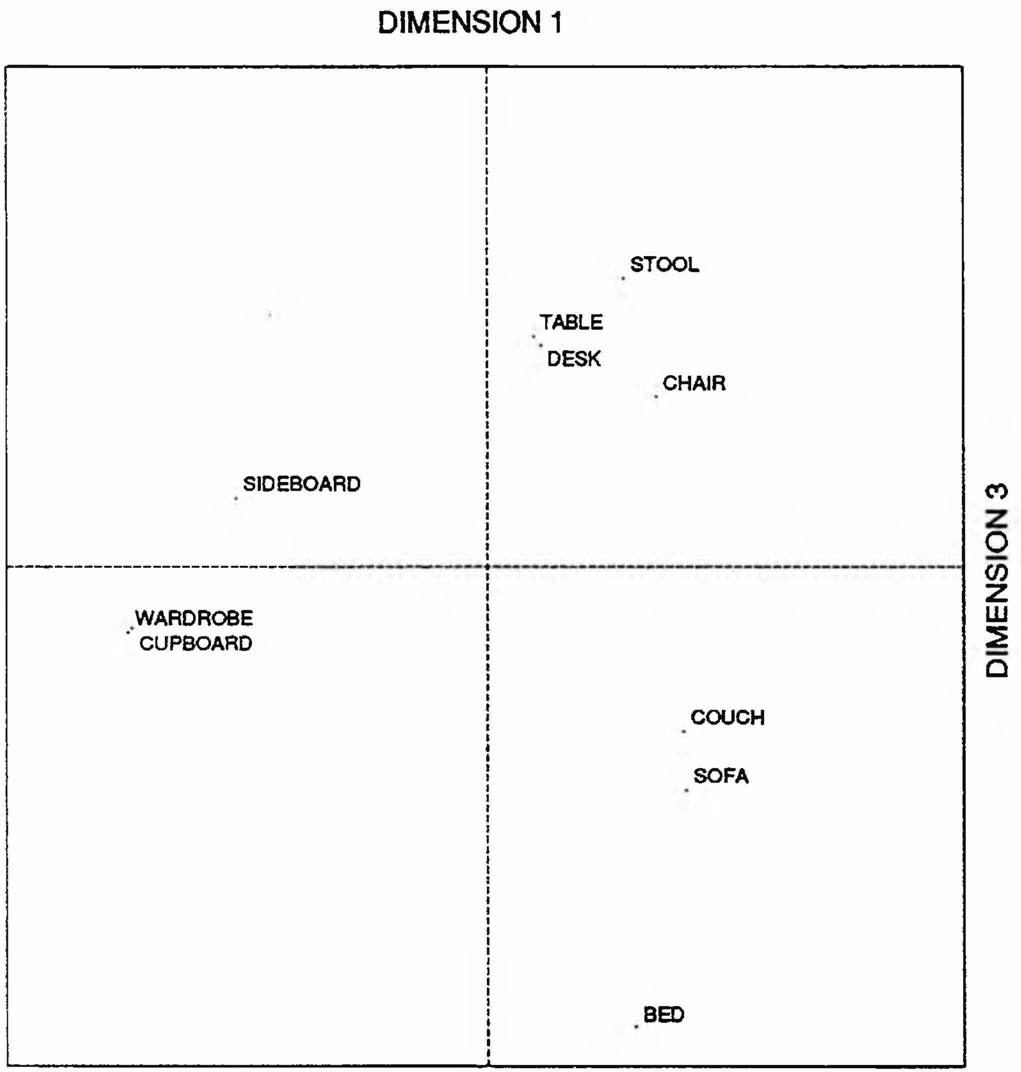


FIGURE 4.1b

Structure derived from semantic similarity judgements of the furniture terms

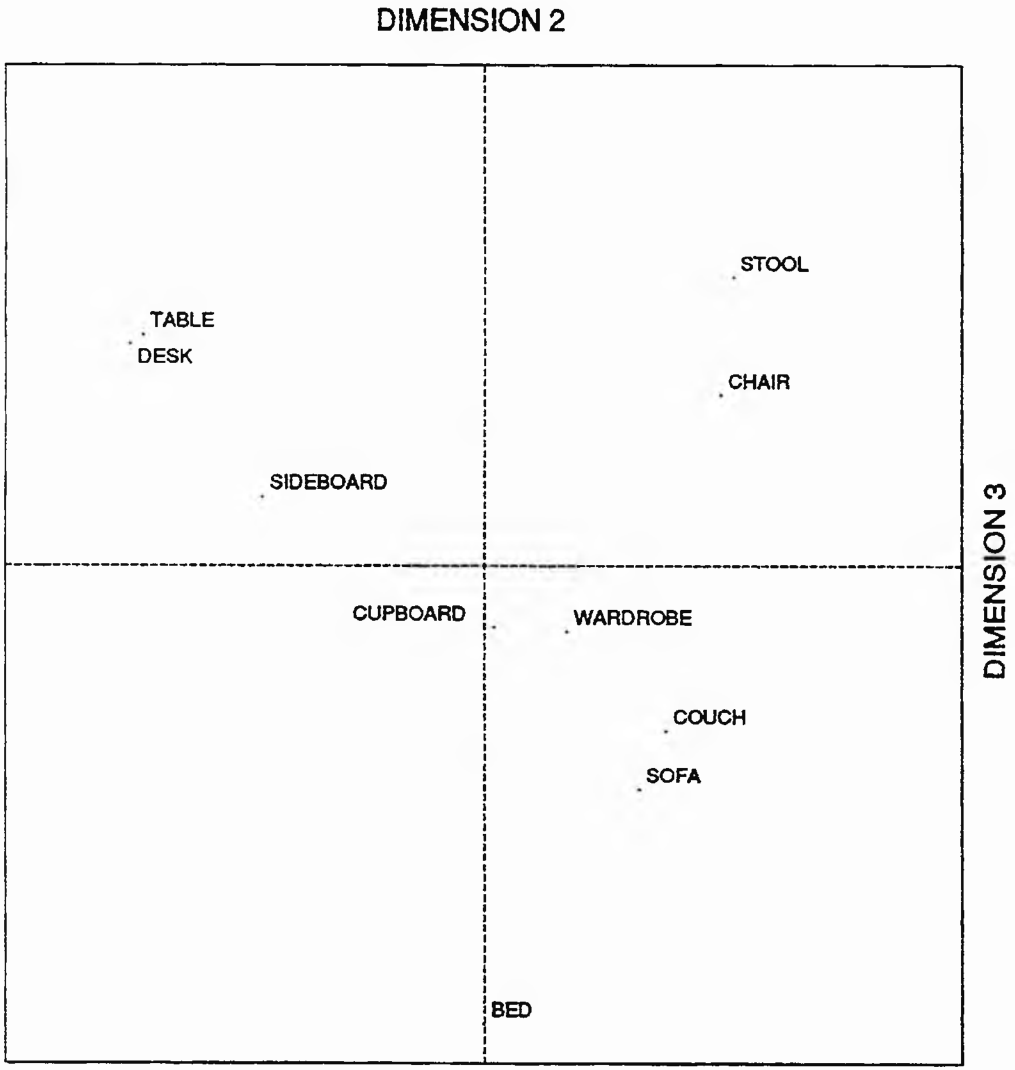


FIGURE 4.1c

Figure 4.1b plots dimension 1 against dimension 3. It is possible that dimension 3 refers to a 'hard' to 'soft' dimension. There is a group of terms in the top half of the figure which are relatively hard (STOOL, TABLE, DESK and CHAIR) which are contrasted with BED which is placed at the very bottom of the plot. COUCH and SOFA are toward the bottom of the figure.

Figure 4.1c plots dimension 2 against dimension 3. A possible interpretation of dimension 2 is that it reflects a dimension related to the degree to which the furniture terms are 'associated with the person' (e.g. CHAIR, STOOL, COUCH) as opposed to 'associated with things' (e.g. DESK, TABLE, SIDEBOARD).

The coordinates of the MDS model depicted in Figures 4.1a to 4.1c were abstracted and employed as the 'semantic similarity' linguistic factor model against which the global ERP variance models were rotated. The numerical values of the model are appended at Appendix 3.

A 3-dimensional MDS model, based on the judgements of orthographic similarity of all the (included) participants in this and the next experiment (N = 15), was derived. This solution explained 0.57 the input matrices' variance. This model is summarized in Figure 4.2, which plots dimension 1 against dimension 2.

WARDROBE, CUPBOARD and SIDEBOARD, as for the semantic model, are clustered together, reflecting the fact that they are all long words, composed of two syllables. Two of these words obviously share the syllable 'BOARD'. The proximity of these items in both the semantic and orthographic models is

Structure derived from judgements of the orthographic similarity of the furniture terms

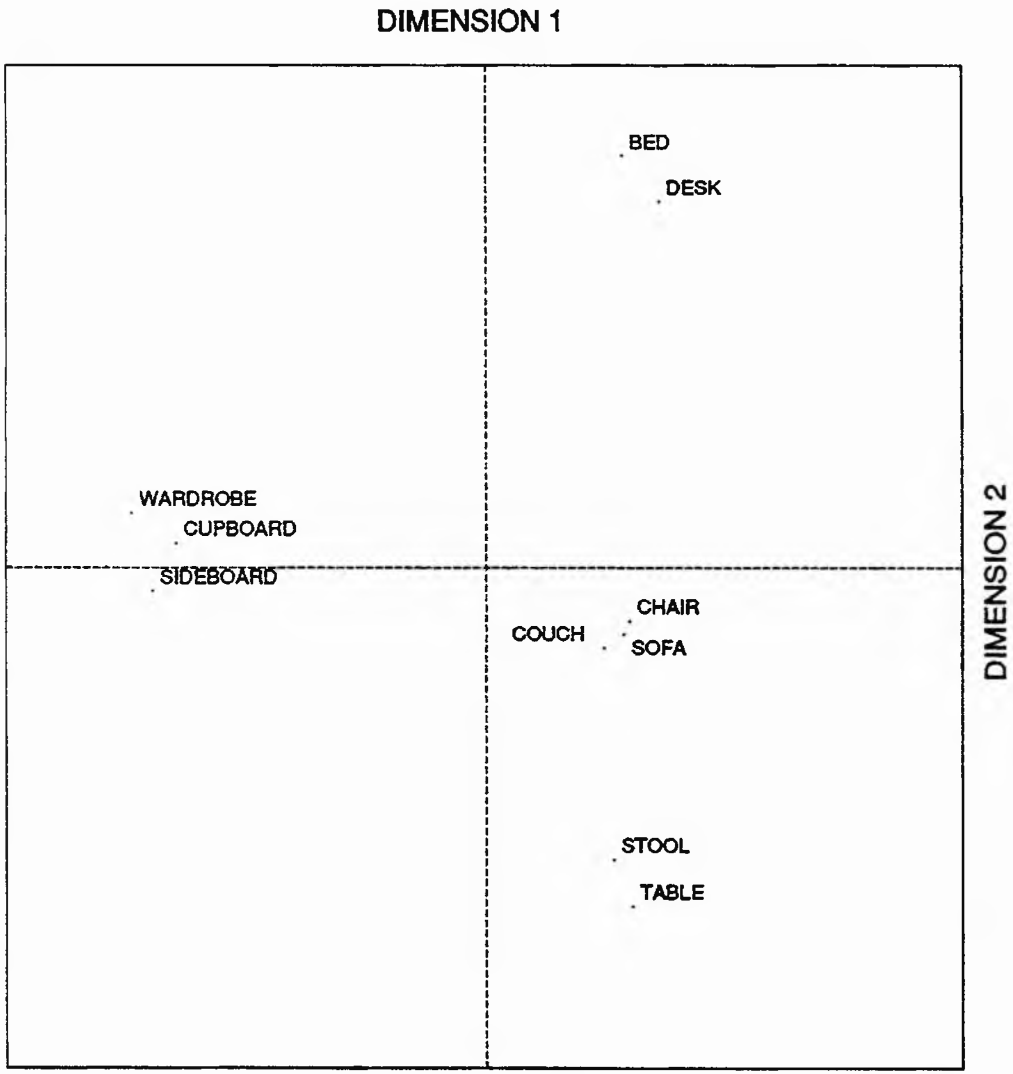


FIGURE 4.2

obviously undesirable: The dissociability of the models is a (negative) function of the degree to which the models share variance.

Another cluster in this 'orthographic' solution is that of BED and DESK, both of which are monosyllabic words with 'E' as their embedded vowel. More puzzling is the cluster exemplarized by COUCH, CHAIR and SOFA. It is not difficult to rationalize the proximity of COUCH and CHAIR: Both are monosyllabic words possessing the 'CH' digram. But the proximity of SOFA to these terms seems less explainable on solely orthographic grounds. It shares the second letter 'O' with COUCH, and an 'A' with CHAIR, but it is suspicious that SOFA, which is a synonym of COUCH, is grouped here. It is possible that semantic factors have intruded into the judgements of orthographic similarity. The proximity of these 3 terms is another source of covariation between the orthographic and semantic models.

The coordinates of the model were employed as the orthographic similarity model and these values are appended in Appendix 3.

The MDS solution based on 15 subjects' judgements of the phonological similarity of the furniture words was again computed in 3 dimensions. This explained 0.549 the input variance. The configuration is summarized by presenting dimension 1 plotted against dimension 2 in Figure 4.3. WARDROBE, CUPBOARD and SIDEBOARD are once more clustered close together. SOFA and STOOL are grouped together, as are COUCH and CHAIR, and DESK and BED. TABLE is somewhat dissociated from the other members

Structure derived from judgements of the phonological similarity of the furniture terms

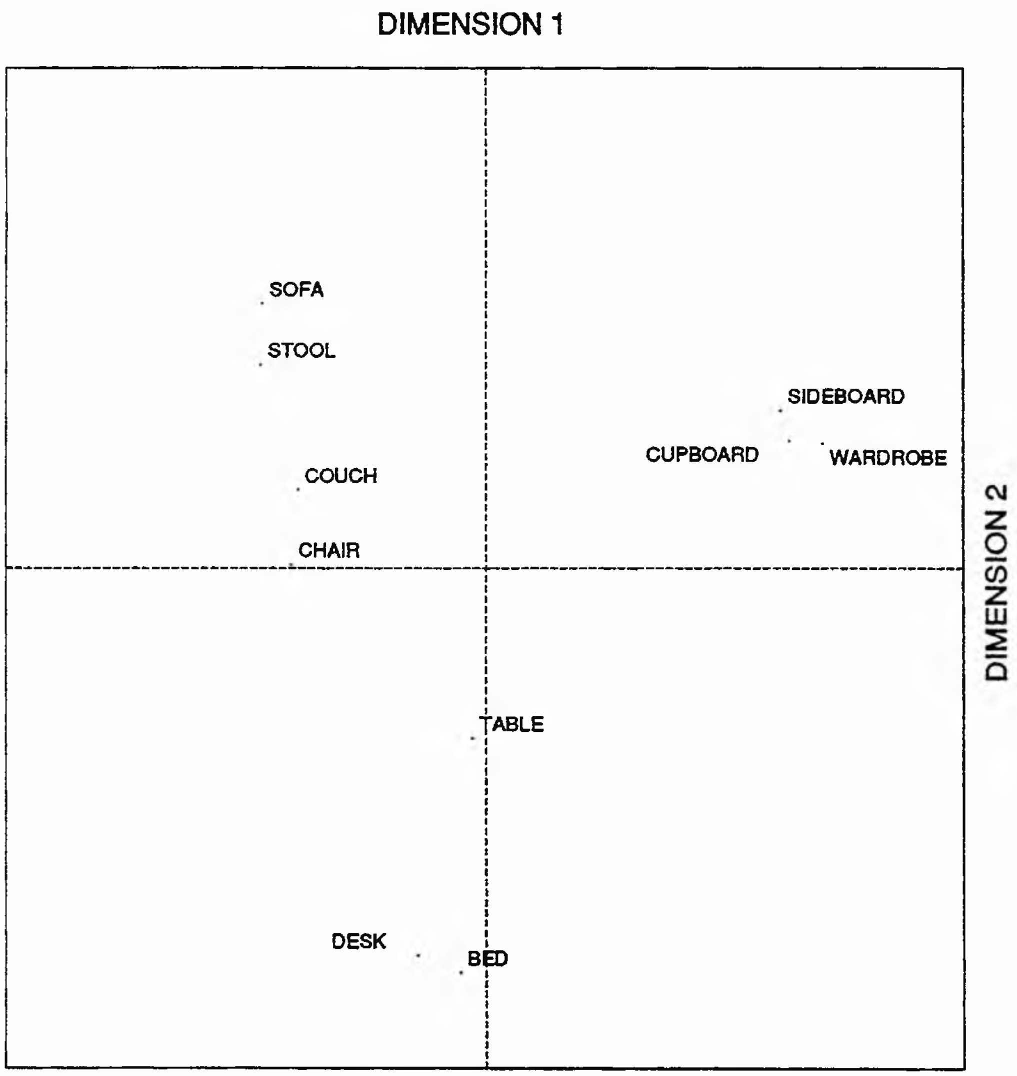


FIGURE 4.3

of the set. As before, the numerical values of the model were abstracted and employed as the 'phonological similarity' model, and the model is appended at Appendix 3.

Interrelations between the linguistic models

A useful way to conceptualize the relations between the global ERP variance models and the linguistic factors which the furniture words embody is to employ a spatial metaphor. For the colour names, I derived a 2-dimensional space which represented the interrelations between the linguistic models for that category. Models which closely covaried were close together in this space. It was possible to identify regions of the space as being related to particular types of linguistic factor (e.g. word frequency). The ERP model could be thought of as being embedded in the plane of the 2-space at a point. The proximity of the point representing the ERP model to the linguistic factor models, and particularly the region of the space which the ERP point occupied, was a guide to the likely factor responsible for 'driving' the ERP variance.

Accordingly, a matrix of proximities was derived for the furniture names by rotating each model onto every other model. This was achieved by PROCRUSTES. The proximity matrix expresses the degree to which the linguistic factor models share variance. This matrix was used as input to an MDS routine (see Chapter 3) and the output configuration expressed the relations between the models in terms of their shared variance. This space is depicted in Figure 4.4.

Interrelations between the linguistic factor models for the furniture terms

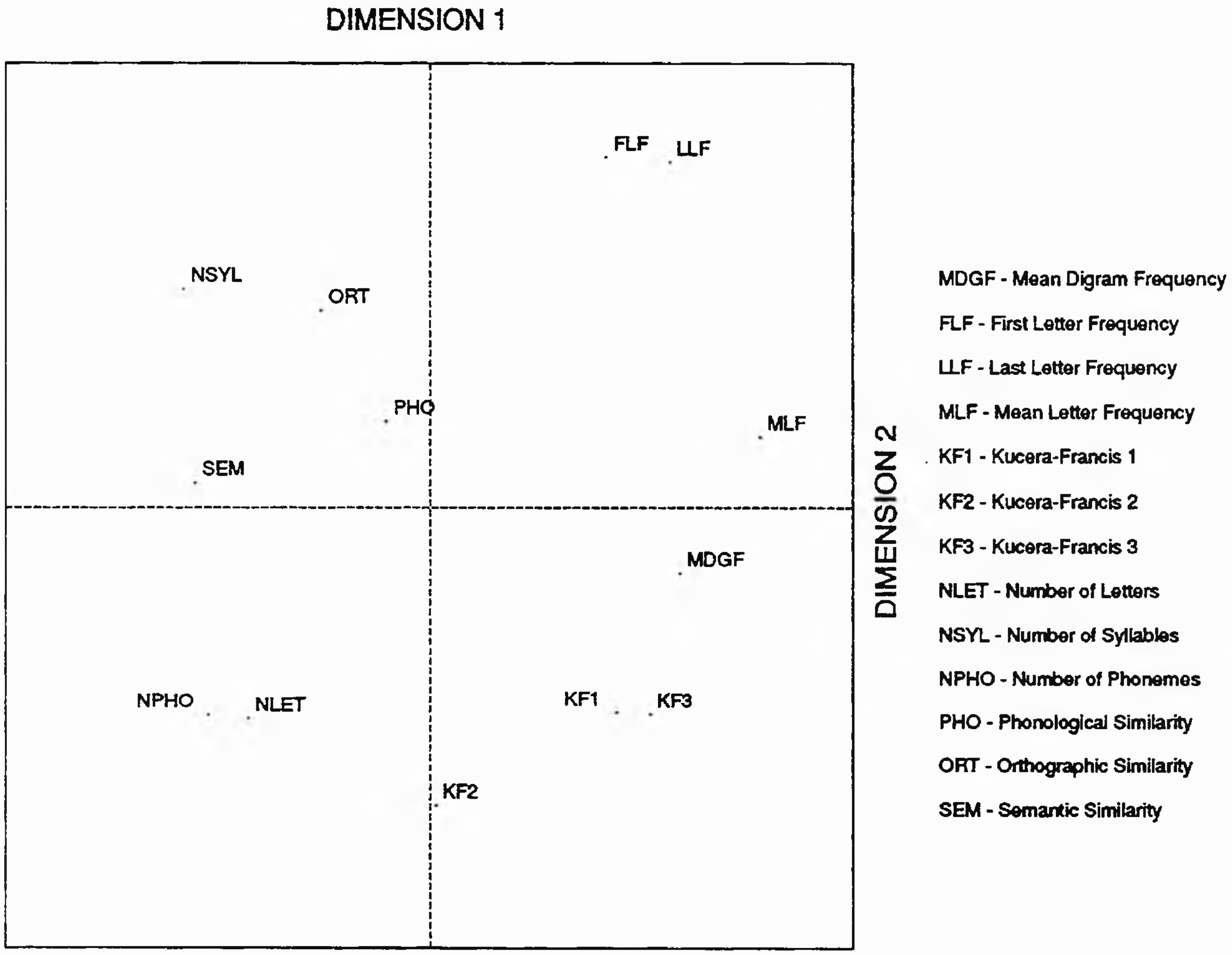


FIGURE 4.4

The upper right region of the figure seems related to the letter frequencies of the words as all the letter frequency models, and no other models, are present. This is in contrast to the 'factor space' derived for the colour names, in which the letter frequency models were distributed almost at random throughout the space (see Figure 3.7). The lower right hand region of the space appears to be related to the word frequency of the words, since all the word frequency models are present.

The lower left region of the factor space seems best understood as related to word length as 'number of letters' and 'number of phonemes' occupy it. The upper right region is occupied by the 3 judgement task models.

The factor space seems, in some respects, better to dissociate the different types of competing linguistic factor models than did the factor space for the colour names. There are 3 broad regions of the space which relate straightforwardly to letter frequency, word frequency and word length. Less positively, the fourth region of the space contains all 3 judgement models, a fact that reflects their close covariation (see above). Their close proximity, and the intuitions gained from conducting the colour name experiments about the level of specificity of covariation between the ERP and linguistic domains, made it seem to me unlikely that the 3 judgement models would be dissociable in the analyses which follow.

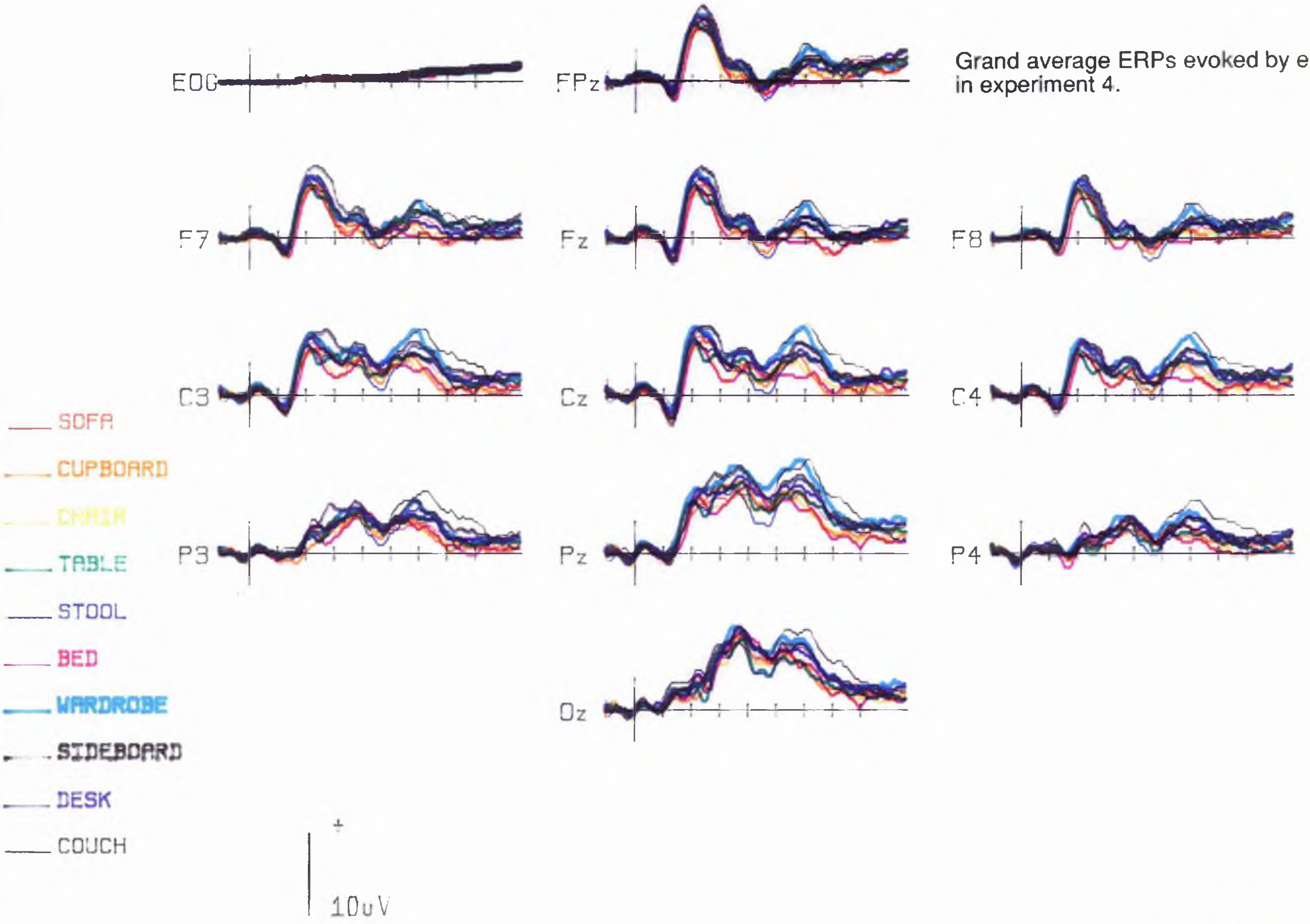
Electrophysiological data

Figure 4.5a and 4.5b show the grand average waveforms from the 7 subjects. The median number of trials comprising each subject's average ERP was 46. The waveform associated with each of the 10 furniture terms is traced in a different colour, which was arbitrarily assigned (see the key at left). The grand average responses did not enter any analysis, and are presented to show the general morphology only. A very similar morphology to that in the colour name category membership decision experiments is present.

Table 4.1 shows the variance-explained statistics for each of the linguistic factor models, when compared to the global ERP variance model computed for data drawn from the whole epoch. 'Kucera-Francis 1' and 'Kucera-Francis 3' are the models whose correspondence to the ERP model is least likely to have come about by chance. 'Digram frequency' and 'phonological similarity' are also associated with relatively low probabilities. 'Digram frequency' is a covariate, for these words, of word frequency (see Figure 4.4), and 'phonological similarity' is the judgement model nearest the word freqency region of the factor space. Hence the ERP model is positioned in the lower right 'word frequency' region of the space, and it may be that, as for experiments 2 and 3, the ERP model for the whole epoch is best related to word frequency.

To preserve comparability with the colour name experiments (1 to 3), early, mid-latency and late time windows were submitted to configurational analysis.

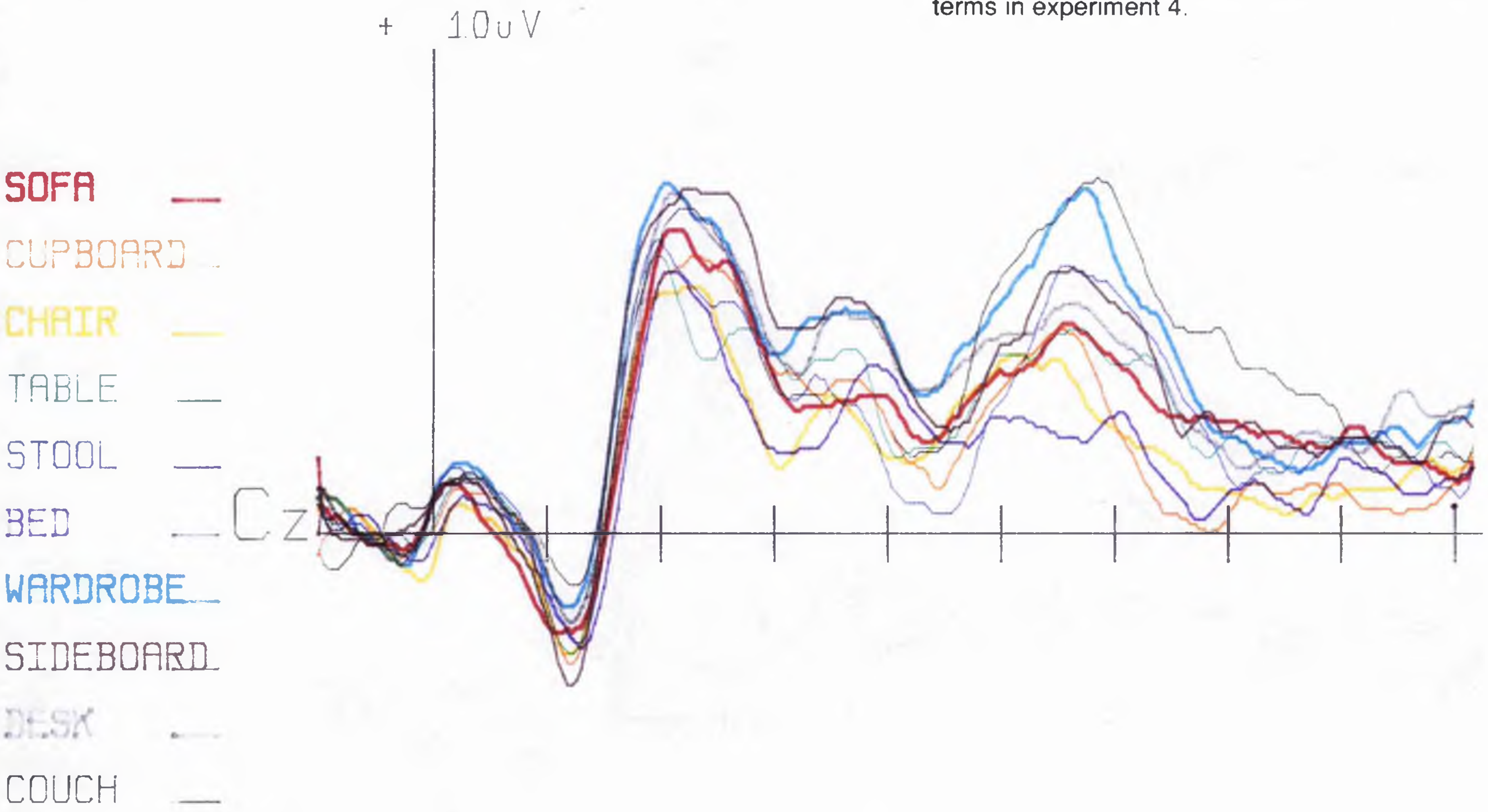
Grand average ERPs evoked by each of 10 furniture terms in experiment 4.



Ticks are at 100 msec intervals

FIGURE 4.5a

Grand average vertex ERPs evoked by each of 10 furniture terms in experiment 4.



Ticks are at 100 msec intervals

FIGURE 4.5b

Table 4.1

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-924 msec (7 subjects in experiment 4).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.191	.14
first letter freq	.066	.68
last letter freq.	.128	.41
mean letter freq.	.059	.75
Word frequency models:		
K-F1	.198	.12
K-F2	.161	.23
K-F3	.202	.11
Word Length models:		
Number of letter	.060	.69
Number of syllables	.104	.43
Number of phonemes	.019	.93
Similarity Judgement models:		
phonological	.321	.19
orthographic	.150	.66
semantic	.151	.65

Early ERP variance

Table 4.2 summarizes the results of the analysis of a time window from stimulus onset to 250 msec. Two of the word length models 'number of syllables' and 'number of letters' are the models least likely to be related to the ERP variance by chance. As for every experiment so far reported, it may be that word length is the most important factor in early ERP variance.

Mid-latency ERP variance

Data drawn from a window extending from 250 msec to 400 msec were analysed by configurational analysis. The results are summarized in Table 4.3. 'Semantic similarity' best explains the ERP model variance, and is associated with the lowest probability that its precision of fit could have come about by chance. This is in accord with expectations made on the basis of the colour name experiments. However, the lack of dissociability of the judgement models, which was anticipated above, seems to be reflected here, since the orthographic model is also quite successful in explaining variance.

Also, the 3 word length models are related to the ERP model at a level which is unlikely to have come about by chance, indeed they are better related to the ERP variance in this window than in the early window. A possible explanation for this could be that word length is an effective factor beyond 250 msec. The 250 msec 'cut' could be too early for these words, and word length-related

Table 4.2

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-250 msec (7 subjects in experiment 4).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.149	.22
first letter freq.	.156	.28
last letter freq.	.159	.27
mean letter freq.	.089	.58
Word frequency models:		
K-F1	.020	.95
K-F2	.018	.93
K-F3	.011	.98
Word Length models:		
Number of letters	.209	.10
Number of syllables	.219	.09
Number of phonemes	.079	.60
Similarity Judgement models:		
phonological	.330	.21
orthographic	.301	.16
semantic	.154	.65

Table 4.3

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 250-400 msec (7 subjects in experiment 4).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.148	.26
first letter freq.	.136	.36
last letter freq.	.149	.25
mean letter freq.	.078	.64
Word frequency models:		
K-F1	.061	.78
K-F2	.123	.38
K-F3	.063	.73
Word Length models:		
Number of letters	.228	.08
Number of syllables	.256	.06
Number of phonemes	.230	.08
Similarity Judgement models:		
phonological	.336	.20
orthographic	.418	.07
semantic	.473	.04

Table 4.4

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 400-800 msec (7 subjects in experiment 4).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.194	.13
first letter freq	.071	.63
last letter freq.	.142	.27
mean letter freq.	.171	.20
Word frequency models:		
K-F1	.287	.03
K-F2	.170	.19
K-F3	.279	.03
Word Length models:		
Number of letters	.121	.36
Number of syllables	.027	.85
Number of phonemes	.018	.92
Similarity Judgement models:		
phonological	.259	.31
orthographic	.163	.55
semantic	.153	.59

variance might therefore not be parcelled out exclusively to the early window (see Chapters 5 and 6, for a detailed treatment of this issue).

Late ERP variance

As shown in Table 4.4, the global ERP variance model derived for the late window, of 400 to 800 msec, was best fitted by word frequency variables. Both 'Kucera-Francis 1' and 'Kucera-Francis 3' are related to the ERP model at a level which would be expected only 3 times in a hundred by chance. This rarity of correspondence is much greater than that associated with any other linguistic factor model.

Discussion of experiment 4

Word length variables explained variance best early in the epoch. The semantic similarity judgement model explained variance best in the mid-latency analysis. Word frequency variables were the best predictors of the ERP variance in the late window. The great similarity between the pattern of results derived from experiment 4 and that for the colour name experiments - remembering that the **actual models** are quite different in this experiment than in the colour name experiments - implies that important aspects of these exploratory studies are replicable. The same 3 linguistic factors appear to relate to the ERP differences as were inferred from the colour name experiments, and these factors do so in the same temporal sequence.

While the word length variables were the best fitted models to the 1-250 msec variance, they were better fitted to variance in the 250-400 msec window. This suggested that the 250 msec 'cut' was probably too early to concentrate word length related variance into the early window. By the same token, however, this feature suggested 2 beneficial aspects of these exploratory analyses. First, it suggests that the latency at which word length effects are present is toward the end of the 1-250 msec window, rather than much earlier (see Chapters 5 and 6). Second it suggests that, in some circumstances, it may be possible that the analytic procedures are capable of signalling the presence of coactive factors; that is, factors whose effects on the ERP overlap in time (see Chapters 5 and 8).

Experiment 5

Method

Subjects: Eight young adults were payed for participation. These 8 subjects were a subset of the odd numbered subjects who took part in the furniture term sessions. One subject was rejected from the data set on the ground that she was subsequently discovered not to have learnt English as her first language. Of the 8 subjects, 7 were female. All reported themselves to be right handed, to have learnt English as their first language and to be free of any prescribed medication. Four subjects had taken part in an evoked potential study before. All were naive as to the purpose of the experiment.

This experiment was in all other respects identical to experiment 4.

Results and discussion

Behavioural data

Reaction time and error rate: The mean reaction time to the out of category targets was 673 msec. The standard deviation of the reaction time distribution was 59 msec. The miss rate was 1.1% and the false alarm rate 0.3%.

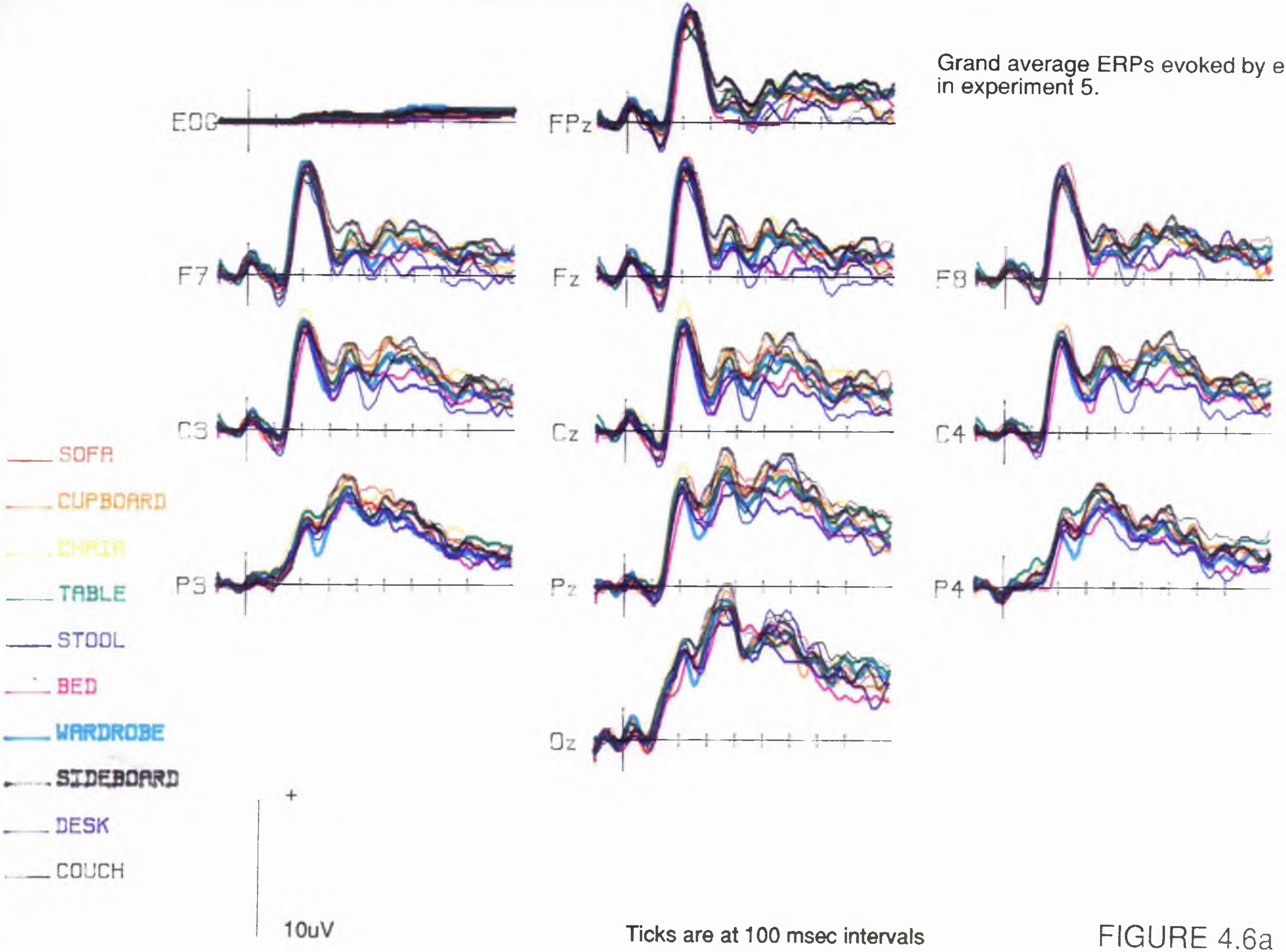
Similarity judgement data: The judgement data collected in this data set were pooled with the data collected in experiment 4. The pooled data were used to derive the semantic, orthographic and phonological configurations described in experiment 4.

Electrophysiological data

Figure 4.6a and 4.6b show the grand average waveforms from the 8 subjects. The median number of trials comprising each subject's average ERP was 47. The waveform associated with each of the 10 furniture terms is traced in a different colour, which was assigned according to the arbitrary scheme chosen for experiment 4 (see the key at left). The grand average responses are presented to show the general morphology of the responses. A very similar morphology is present to those in experiment 4 and in the colour name category membership decision experiments.

In the now familiar way, the first configurational analysis was based on data drawn from a time window extending the entire length of the post-stimulus epoch. Table 4.5 shows the variance-explained statistics for all the linguistic factor models against the ERP model. There is scant evidence for the effects of any specific model or model group. The word frequency models are among the least likely models to be related to the ERP variance by chance (but note that the phonological similarity model is, in fact, the least likely), but no models 'stick out' from their competitors in this analysis.

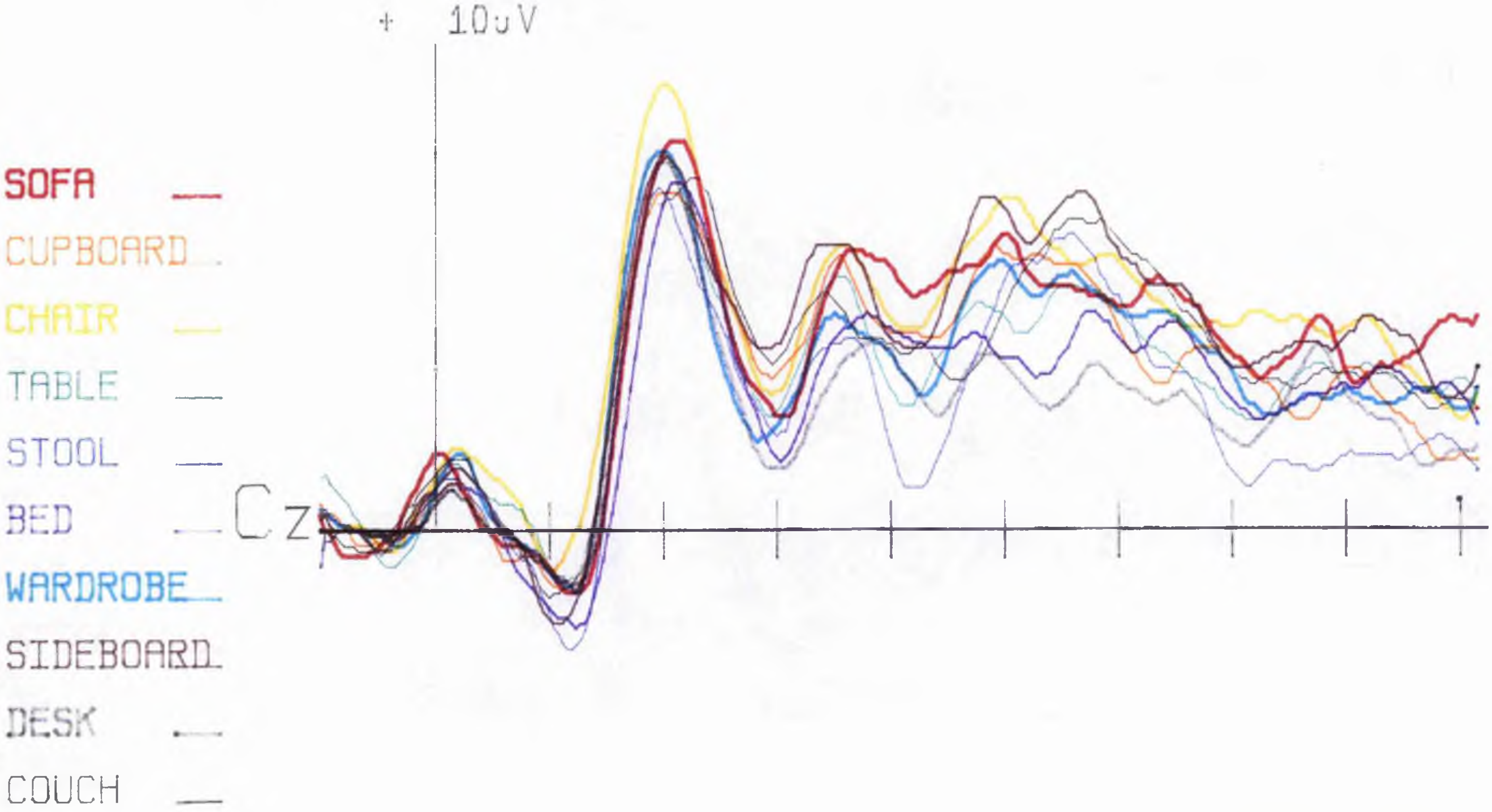
Grand average ERPs evoked by each of 10 furniture terms in experiment 5.



Ticks are at 100 msec intervals

FIGURE 4.6a

Grand average vertex ERPs evoked by each of 10 furniture terms in experiment 5.



Ticks are at 100 msec intervals

FIGURE 4.6b

Table 4.5

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-924 msec (8 subjects in experiment 5).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.081	.52
first letter freq.	.004	.99
last letter freq.	.160	.25
mean letter freq.	.018	.92
Word frequency models:		
K-F1	.118	.32
K-F2	.115	.35
K-F3	.129	.30
Word Length models:		
Number of letters	.111	.44
Number of syllables	.084	.48
Number of phonemes	.028	.87
Similarity Judgement models:		
phonological	.278	.27
orthographic	.174	.54
semantic	.158	.58

Early ERP variance

Variance-explained values for each linguistic model, and the probabilities associated with each, are presented in Table 4.6. There is no evidence whatsoever for the effects of any linguistic variable on ERP variance in this earliest latency window.

Mid-latency ERP variance

As in every experiment, the semantic similarity judgement model fits best in the 250-400 window, as is shown in Table 4.7. 'Semantic similarity' explains slightly more than half the ERP variance, a correspondence which would be expected only once in a hundred permutations by chance. The anticipated problem with dissociating the judgement models is apparent, as it was in experiment 4, however, as both the other judgement models explain susbstantial amounts of variance. Indeed, 'orthographic similarity' is associated with a probability as low as that for the semantic model. Whether it is reasonable to attribute the high variance-explained statistics for these competing judgement models to their close covariation with, and hence close proximity in the 'factor space' to, the semantic model is addressed in Chapter 5.

Another feature of these results, which was very similar to those of experiment 4, was that the word length models explained variance to a greater extent in this mid-latency window, and did so to a degree which was statistically rare. This, again, could be taken to imply that the 250 msec cusp (which was retained for

Table 4.6

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 1-250 msec (8 subjects in experiment 5).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.064	.75
first letter freq.	.038	.85
last letter freq.	.034	.88
mean letter freq.	.104	.49
Word frequency models:		
K-F1	.102	.51
K-F2	.112	.44
K-F3	.118	.42
T-L		
Word Length models:		
Number of letters	.119	.45
Number of syllables	.065	.63
Number of phonemes	.050	.78
Similarity Judgement models:		
phonological	.158	.80
orthographic	.113	.85
semantic	.154	.67

Table 4.7

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 250-400 msec (8 subjects in experiment 5).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.151	.22
first letter freq.	.101	.55
last letter freq.	.168	.19
mean letter freq.	.073	.65
Word frequency models:		
K-F1	.042	.81
K-F2	.046	.75
K-F3	.033	.87
Word Length models:		
Number of letters	.276	.04
Number of syllables	.052	.69
Number of phonemes	.218	.10
Similarity Judgement models:		
phonological	.315	.11
orthographic	.493	.01
semantic	.507	.01

consistency with the colour name experiments) was too early for these words, and that variance sensitive to word length was present here but was distributed later in time. The more fine grained aspects of the distribution of word length effects in time are treated in Chapters 5 and 6.

Late ERP variance

The results of a configurational analysis undertaken on data drawn from a latency window from 400 to 800 msec are shown in Table 4.8. The two models least likely to have explained the ERP variance by chance are 'Kucera-Francis 1' and 'Kucera-Francis 3'. Again, word frequency variables tend to explain variance best in the 400-800 msec window, though these results, if taken in isolation from the pattern of results over the 4 category decision data sets explored so far, could easily be due to chance.

Table 4.8

Variance-explained values and associated probabilities for each linguistic model against the ERP global variance model for 400-800 msec (8 subjects in experiment 5).

MODEL	R^2	p
Letter frequency models:		
digram frequency	.149	.23
first letter freq.	.078	.60
last letter freq.	.104	.49
mean letter freq.	.065	.68
Word frequency models:		
K-F1	.172	.19
K-F2	.129	.31
K-F3	.161	.21
Word Length models:		
Number of letters	.036	.82
Number of syllables	.135	.30
Number of phonemes	.036	.79
Similarity Judgement models:		
phonological	.311	.23
orthographic	.121	.77
semantic	.078	.91

Discussion of experiment 5

My interpretation of the results of these analyses was that, with the possible exception of word frequency, the same linguistic factors as have been repeatedly identified appeared to be affecting ERP variance in this experiment.

The semantic similarity model was again better fitted to the ERP variance in the mid-latency analysis than was any other competing model. Similarly this good fit was confined to the mid-latency analysis, in that semantic similarity did not explain variance well in the other time windows which were subjected to analysis. The high variance-explained values for the 2 other judgement models could be interpreted as due to their close covariation with the semantic model. But a clearer conclusion on this issue requires that the pattern of results over the whole of Part One be examined. As noted at the end of Chapter 3, any one small set of words may possess adventitious dependencies between the linguistic factors which it embodies, such as those between the judgement models for the furniture words. The same relations did not exist for the colour names, in whose factor space the judgement models were dissociable (see Chapter 5).

With similar consistency, the word length variables explained variance best in the earlier parts of the epoch. But, more clearly in these data than in the previous experiment, the latencies at which the window most sensitive to word length should begin and end seem to be later than the 1 to 250 msec partition in these analyses (see Chapters 5 and 6).

General Discussion

Several methodological aspects of the above furniture term experiments were different to the colour word experiments. First, the individual words which were employed to evoke the ERPs were obviously quite different to the colour words. Second, the numerical values of the linguistic factor models, against which the furniture word ERP models were compared, were completely different to those of the linguistic factor models which were compared to the colour word ERP models. Third, a different electrode montage was employed in the furniture word experiments, which was referred to a different reference electrode pair.

Despite these differences, it appeared that the same linguistic factors; word length, semantic similarity and word frequency were identifiable as the factors most likely to be responsible for the ERP differences. Single word ERPs, evoked by exposure to furniture words, differed in the same ways as single word ERPs evoked by exposure to colour names. In both categories the differences between the ERPs were arranged according to the orders expected by the same linguistic factors.

This underlying consistency was clearer in experiment 4. The best predictors of ERP variance in the earliest window of experiment 4 were the word length models. These were followed by the semantic similarity model in the mid-latency analysis. Word frequency models best explained ERP variance in the late window. This temporal sequence was qualitatively identical to that for experiment 2. The only qualitative difference was that the word length effects appeared to be later in time for the furniture words than for the colour names.

This later distribution of the word length effects may have been more apparent in experiment 5. The word length models did not explain variance well in the earliest window, but were effective predictors in the mid-latency analysis. Taken without the context of experiment 4, this could be viewed as an indication that a different, and therefore inconsistent, temporal order in the efficacy of the linguistic factors was apparent: That word length and a semantic factor were coactive in the mid-latency region of the epoch. But, in the context of the suggestion that word length effects were distributed somewhat later in time in these experiments, a conclusion derived from experiment 4, it seemed likely that the same phenomena were being observed. That is, that a word length effect was followed by an effect due to semantic similarity but that the chosen time windows were too broad (and badly placed) to partition the effects in the case of experiment 5. It is certainly not the case that the reverse temporal order of these factors was apparent. In order to further substantiate the intuition that experiment 5 was consistent with the results of the foregoing experiments it is necessary to undertake temporally more fine-grained analyses. These analyses, for all the data sets, are reported and discussed in Chapter 5.

In both experiments 4 and 5, the semantic model related well to the ERP differences in the mid-latency analyses. In both cases, the precision of fit was such that it would be unlikely to come about by chance. It is still less likely that both experiments would, purely by chance, identify this factor as the most important one (of 13) in the same latency region (of 4). Hence, these experiments provide evidence that a semantic factor may underlie mid-latency variance in ERPs evoked by individual words in a category membership decision task. Inference from these results, however, is complicated by the close covariance of the semantic model with both the phonological and orthographic models. As remarked above, the influence of these factors may only be disambiguated by considering the patterns of variance explanation over all 5 quantitative modelling experiments. This will be accomplished in Chapter 5.

Finally, while the evidence for the influence of word frequency on the late window of experiment 4 was strong, the comparable analysis in experiment 5 provided only weak, but not inconsistent, evidence for the influence of this factor. The likelihood that this factor is effective on the waveform at this latency is evaluated, by examining all the data sets, in Chapter 5.

Chapter Five

Summary of Part One

Introduction

The analyses carried out on the 5 data sets which constitute Part One raise a number of issues which I will attempt to address in this Chapter. In particular, I am going to examine the questions of which linguistic factors are most strongly implicated, what the time course of the sensitivity of ERP variance to each of the likely factors might be, and what features of the ERPs mediate the likely factors. The rationale for integrating the results presented in the foregoing two chapters is that some aspects may only be disambiguated by examining the pattern of results over all the experiments.

Which linguistic factors underlie ERP variance in these paradigms?

This question is, of course, the central question with which the whole enterprise is concerned (see Chapter 1). Some of the experiments, particularly experiments 1, 2 and 4, produced results which were relatively clear, in that, in some latency windows, they identified only one linguistic factor as being well fitted to the empirical variance, at a level of precision which was unlikely to have come about by chance. Some of the experiments, particularly experiments 3 and 5, were more equivocal in some respects. There are problems, however, associated with laying too much emphasis on the results of any one experiment. Before integrating the results of all five sets of analyses and converging on the most probable factors, these problems will be addressed.

The first problem is that, in any one experiment, there may have been sequence effects conflated with the effects due to any of the linguistic factors (see below).

The second is that for any one set of words there are adventitious dependencies, reflected in high covariance, between some of the linguistic factor models. These dependencies could give rise to high variance-explained statistics for some models simply by virtue of their shared variance with the model which is, in fact, responsible for the ERP differences (see below).

The third problem is that, as detailed in Chapter 2, relatively less emphasis should be placed on individual p-values, and relatively more emphasis placed on consistency across samples (Muller et al 1983). This is particularly the case in exploratory studies conducted in an innovative fashion in a domain (single word ERPs) where there are no empirical precedents by which to be guided. This aspect has already been discussed in Chapter 2.

The fourth problem is that the latency points at which the epoch was partitioned may not have been positioned optimally. If the time windows are segregated at a point half way, say, through the time course of the influence of a linguistic factor on the waveforms, variance related to this factor may be insufficiently great in either window to be identifiable, given the presence either of noise or of another factor (see below).

Sequence effects

It is known that the relations between an item and the item (or items) which precede it can give rise to ERP variance. This can be brought about in at least two ways. The first possible source of sequence effects lies in differences in the subjective probability of a stimulus given its precedents (Duncan-Johnson and Donchin 1977). The concomitant of this type of sequence effect is variance associated with the P300 component of the ERP waveform (Duncan-Johnson and Donchin 1977).

A second possible source of sequence effects is due to associative relations obtaining between successive items (Bentin, McCarthy and Wood 1985). The electrophysiological concomitant of this type of sequence effect is variance associated with the N400 component of the waveform.

The sequences of stimuli in experiments 1 to 5 were random, in the sense that there was a constant probability that any item might follow any other. But this random property was ensured by the choice of the particular stimulus on any trial being determined by a random-number table. This table was the same for all subjects in any one experiment.

The consequence of the use of the same random-number table within an experiment was that sequence effects may have been conflated with variance related to the linguistic factors in any one experiment. This was a defect of the experimental designs employed in Part One (which is not repeated in Part

Two). Different random-number tables, however, were employed across experiments. Sequence effects, if present, would therefore be expected to take a different character in the different experiments.

The possible problem of sequence effects can, therefore, be addressed by looking at the pattern of which models are associated with high variance-explained statistics across experiments (experiment 1 had a different table to experiments 2 and 3, which in turn had a different table from that of experiments 4 and 5). It is very unlikely that the same structure of variance could be generated across the experiments by any sequence effects which may be present. Consistencies between experiments in the models which best explain variance in particular latency windows cannot be attributed to sequence effects.

These considerations imply that sequence effects, even in the very unlikely circumstance that they give rise to variance which may be misattributed to one of the linguistic factors, can be disambiguated by examining the results across experiments.

In passing, an account of the results of these experiments, given in terms of the action of sequence effects alone, would face a number of other difficult questions. Why should random sequence effects give rise to variance associated with word frequency in the latency range of P300? Why should the same random sequence effects give rise to variance associated with semantic similarity in the latency region of N400? Why should sequence effects affect variance in the latency region of P200 at all, let alone endow it with a structure similar to that expected if it were due to word length? I can find no reasons to suppose that sequence effects could have given rise to these apparent correspondences with the different linguistic factor models at different latencies. I consequently suspect that, if present, sequence effects would have contributed to noise and therefore to the general difficulty of fitting systematic variance.

Dependencies between linguistic factor models

As remarked several times, the linguistic factors which words embody are not orthogonal, nor are the models which are intended to capture them in these analyses. Interpreting the variance-explained statistics associated with particular models, therefore, is complicated. In seeking to disambiguate covarying factors, I have avoided the practice of orthogonalizing the factors by application of PCA (cf., Whaley 1978), for the reasons detailed in Chapter 2, and have instead relied on a knowledge of the relations between the factors. This knowledge of the interrelations between the factors was generated by computing all-pairwise shared-variance statistics between the factor models and expressing the resultant proximity matrix in a 'factor space' derived by MDS (e.g. Figure 3.7 and Figure 4.4).

The computed factor spaces are different for the two groups of words. For example, in the factor space derived for the colour names, 'orthographic similarity' is revealed to be a close covariate of word length. In the factor space derived for the furniture terms, the same model is revealed to be a covariate of the other judgement models, particularly the strongly implicated 'semantic similarity' model. Similarly the 'first letter frequency' model was a covariate of word length for the colour names, while for the furniture terms it was not.

These kinds of differences between the two factor spaces can be exploited to provide a means of disambiguation. Particularly, these differences are useful in clarifying which models are more likely to have been the effective ones in the case that 2 or more models are well related to an ERP model.

For example, 'orthographic similarity' and 'semantic similarity' covaried for the furniture terms. Symptomatic of their covariance was the finding that both models were related to the ERP variance in the mid-latency analyses both of experiment 4 and experiment 5, at a level of precision unlikely by chance. But there was no indication that 'orthographic similarity' explained variance well in the corresponding windows of the colour name experiments, while there certainly was an indication that 'semantic similarity' was an effective predictor of variance at that latency. Hence, it is much more likely that 'semantic similarity' was the important factor at that latency. By the same token, 'orthographic similarity' covaried with word length for the colour names. Symptomatic of this relation was the finding that the orthographic model explained variance well in the early time window of experiment 3, where word length was the expected factor. There was no indication that 'orthographic similarity' explained variance in the comparable early windows of the furniture term experiments, in which the model did not covary with word length.

If ’orthographic similarity’ had been an effective factor in its own right across both sets of words, it would be expected that it would have ’carried’ its covariates with it in both cases. Supposing that the latency window in which ’orthographic similarity’ was active was the 250 to 400 msec window (in which it achieved its lowest probability, in experiment 5), this would have resulted in word length being well fitted in the mid-latency analyses of the colour name experiments. This was not the case. Supposing, obversely, that the orthographic model was effective in the early window (in which it was once the only rare model, in experiment 3), it would have enhanced the ability of the semantic model to explain variance in the early window of the furniture term experiments. This, too, was not the case.

These aspects of the results are not explainable in terms of ’orthographic similarity’ being an effective factor which could ’carry’ its covariates to high values of variance-explained in the windows in which it might have been active. These results, however, *are* explainable in terms of two factors, word length and semantic similarity, ’carrying’ ’orthographic similarity’ in the window in which each was effective, according to the factor with which the orthographic model covaried.

Arguments of similar form can be presented as militating against other candidate linguistic factors, as below.

Another tool, which may be an aid to disambiguating covariations between the models and, therefore, also helpful to the task of identifying the most probable linguistic factors, is to average across the results of the comparable windows of the experiments. I have already remarked (see Chapter 3) that the 3-dimensional judgement models, in general, explain more variance than do the univariate models, but that this advantage is removed by the randomization tests. The derived probabilities therefore provide, quite apart from their role as inferential statistics, a **metric** by which the performance of the models may be assessed.

Consequently, I decided to average across the probability values for each model in the early, mid-latency and late windows of the experiments. I stress that this procedure loses the inferential property of the probabilities and refers to them only as metrics. This process yields an average probability that any model was fitted to the ERP variance in each of 3 latency windows. The average probabilities collapse across possible sequence effects and across the dependencies between models which are private to each set of words. The average probabilities, however, give a guide only to the *relative* probability that any model is related to the ERP variance in each window by chance. Accordingly, the average probability values of each model (excluding 'Thorndike-Lorge' which was not rotated against the furniture term ERPs) are presented on an arbitrary scale, with a dashed line through the mean probability, in Figures 5.1 to 5.3. I shall refer to these figures in the discussion of the candidacy of each linguistic factor model, to which I now turn.

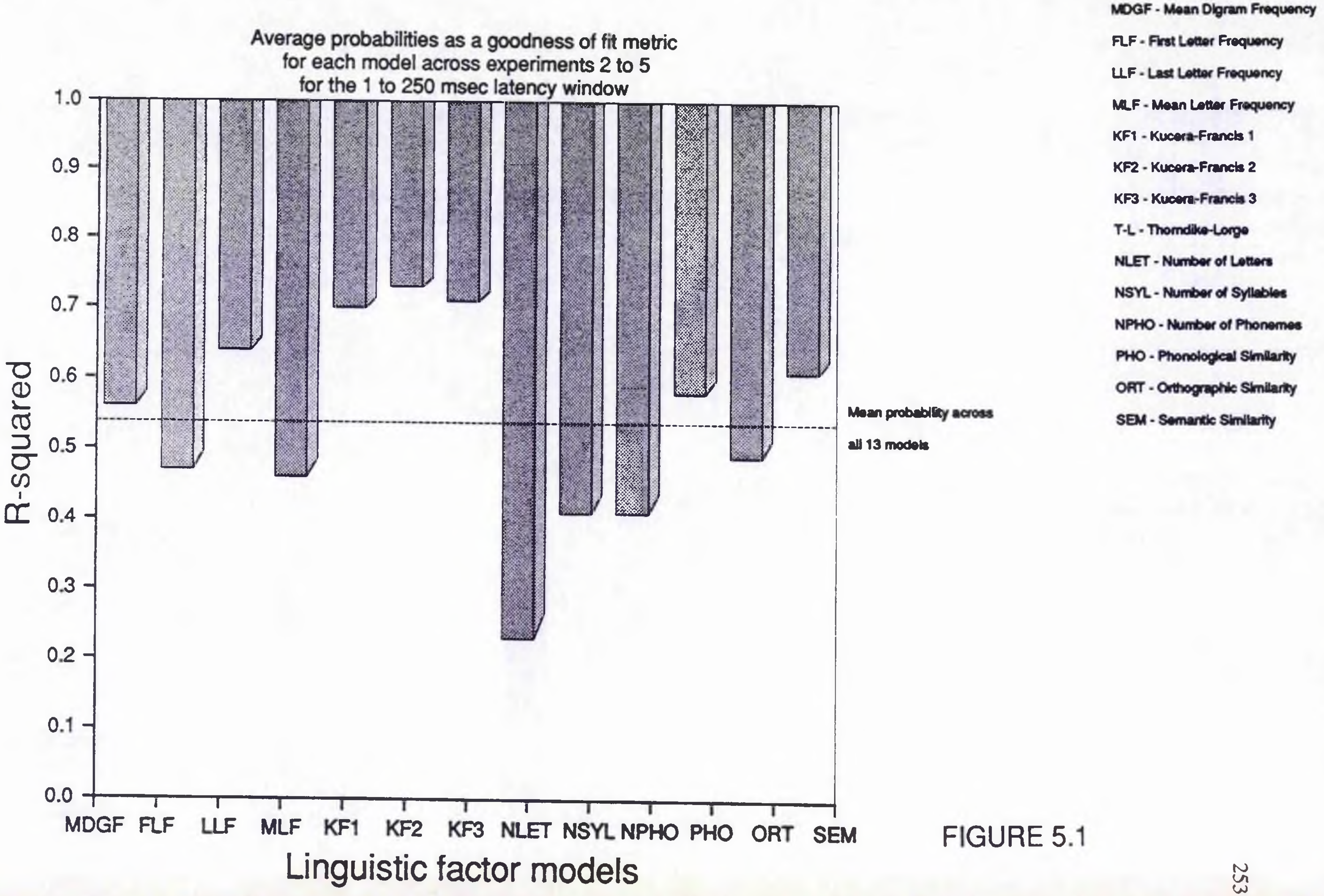

Average probabilities as a goodness of fit metric
for each model across experiments 2 to 5
for the 1 to 250 msec latency window
R-squared
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
MDGF
FLF
LLF
MLF
KF1
KF2
KF3
NLET
NSYL
NPHO
PHO
ORT
SEM
Linguistic factor models
Mean probability across
all 13 models
MDGF - Mean Digram Frequency
FLF - First Letter Frequency
LLF - Last Letter Frequency
MLF - Mean Letter Frequency
KF1 - Kucera-Francis 1
KF2 - Kucera-Francis 2
KF3 - Kucera-Francis 3
T-L - Thorndike-Lorge
NLET - Number of Letters
NSYL - Number of Syllables
NPHO - Number of Phonemes
PHO - Phonological Similarity
ORT - Orthographic Similarity
SEM - Semantic Similarity

FIGURE 5.1

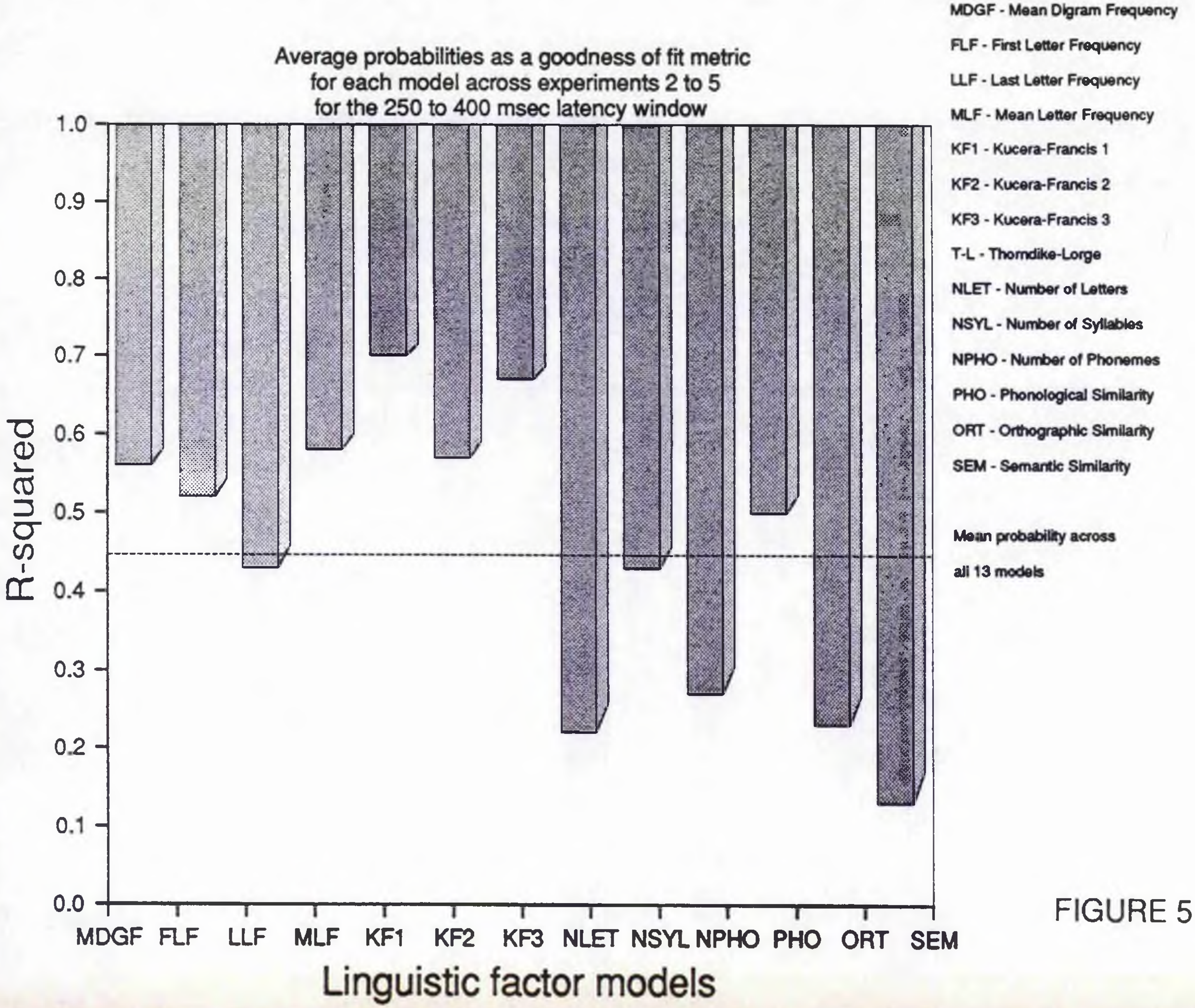

Average probabilities as a goodness of fit metric
for each model across experiments 2 to 5
for the 250 to 400 msec latency window
R-squared
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
MDGF
FLF
LLF
MLF
KF1
KF2
KF3
NLET
NSYL
NPHO
PHO
ORT
SEM
Linguistic factor models
MDGF - Mean Digram Frequency
FLF - First Letter Frequency
LLF - Last Letter Frequency
MLF - Mean Letter Frequency
KF1 - Kucera-Francis 1
KF2 - Kucera-Francis 2
KF3 - Kucera-Francis 3
T-L - Thorndike-Lorge
NLET - Number of Letters
NSYL - Number of Syllables
NPHO - Number of Phonemes
PHO - Phonological Similarity
ORT - Orthographic Similarity
SEM - Semantic Similarity
Mean probability across
all 13 models

FIGURE 5.2

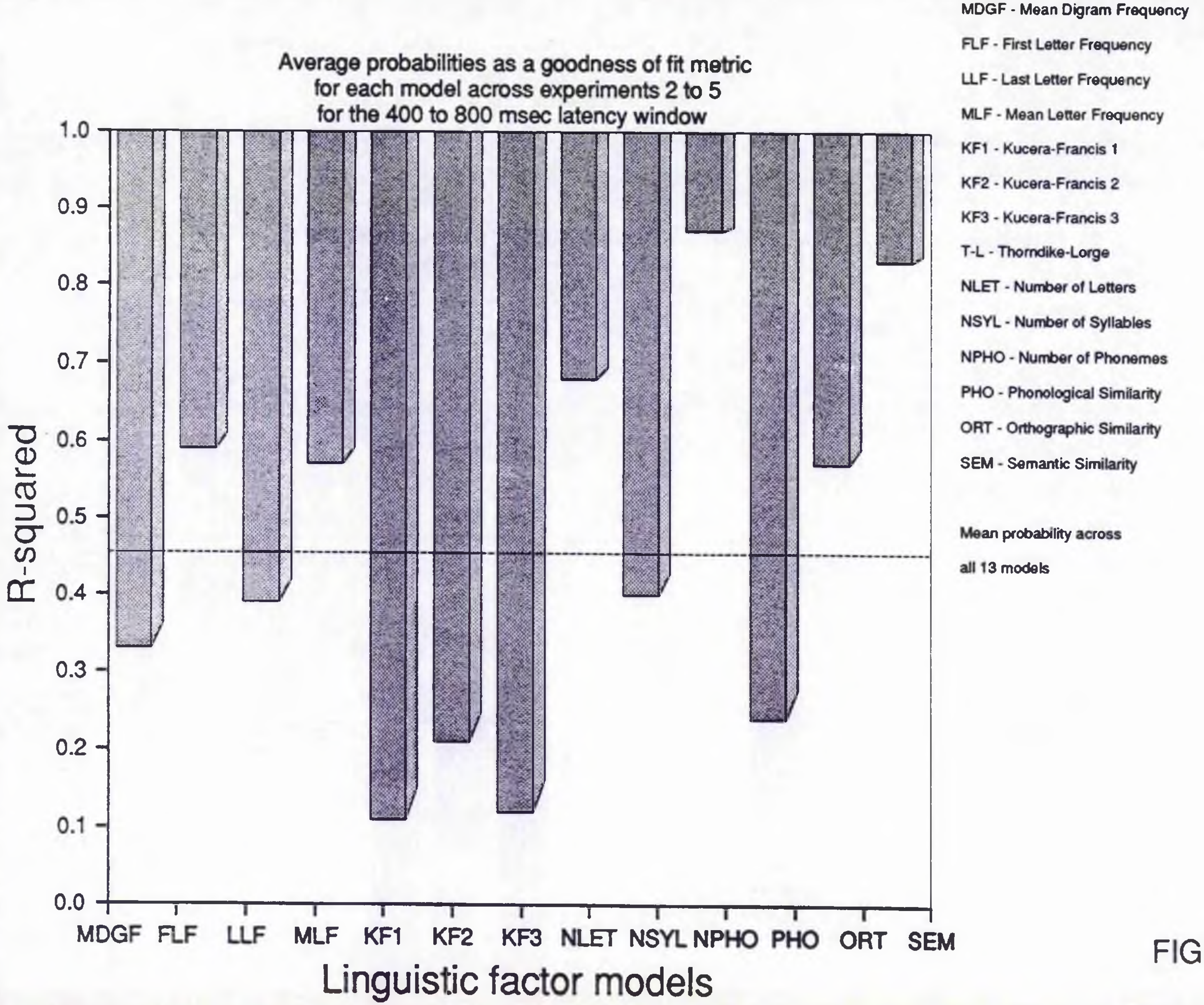

Average probabilities as a goodness of fit metric
for each model across experiments 2 to 5
for the 400 to 800 msec latency window
R-squared
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
MDGF
FLF
LLF
MLF
KF1
KF2
KF3
NLET
NSYL
NPHO
PHO
ORT
SEM
Linguistic factor models
MDGF - Mean Digram Frequency
FLF - First Letter Frequency
LLF - Last Letter Frequency
MLF - Mean Letter Frequency
KF1 - Kucera-Francis 1
KF2 - Kucera-Francis 2
KF3 - Kucera-Francis 3
T-L - Thorndike-Lorge
NLET - Number of Letters
NSYL - Number of Syllables
NPHO - Number of Phonemes
PHO - Phonological Similarity
ORT - Orthographic Similarity
SEM - Semantic Similarity
Mean probability across
all 13 models

FIGURE 5.3

On the candidacy of each linguistic factor model

Mean digram frequency: This model never explained variance in the global ERP variance models to a level which was statistically rare. The lowest probability with which it was associated was 0.10, in the early latency window of experiment 2. This correspondence was not repeated in any of the early windows of the other experiments, indeed this model explained no variance at all in the comparable window of experiment 3.

The only window in which the mean digram frequency model was below the average probability across experiments was the 400 to 800 window (see Figure 5.3). This probably reflected the fact that this model covaried with word frequency for both sets of words (see Figures 3.7 and 4.4).

There was no evidence, therefore, that 'mean digram frequency' was an important factor underlying ERP variance.

First letter frequency: This model explained variance in the early window of experiment 1 to a probability level of 0.09. It was, for a time, considered a possible candidate linguistic factor. This level of improbability, however, was not repeated in any other early analysis, and this model's average probability across experiments for this early window was 0.47, only slightly below the average probability and well away from the value for the most likely candidate model.

The model did not explain variance in any other window of any experiment to a level which was unlikely by chance. The average probability associated with this model in the mid-latency and late analyses (see Figures 5.2 and 5.3) was above that for the average for all the models. It seems unlikely that this model was a predictor of ERP variance.

Last letter frequency: This model never explained variance to a degree that was unlikely by chance and was not associated with probabilities across experiments below the average for the models in general in any window. It was not, therefore, a likely candidate linguistic factor model.

Mean letter frequency: 'Mean letter frequency' was associated with a relatively low probability in the 1 to 250 msec analysis of experiment 2. This was, however, not repeated in other comparable analyses. The average probability over the experiments for this model in this window was only just below that for the average of the models as a whole (see Figure 5.1).

The model gave above average probability values across experiments in the other time windows. No case can be made that this linguistic factor was responsible for the ERP variance, therefore.

Kucera-Francis 1: This model was associated with very low probabilities that its goodness of fit with the global ERP models could have come about by chance in 3 time windows of 2 experiments. These windows were the whole epoch and 400 to 800 msec analyses of experiment 2, and the 400 to 800 msec analysis of experiment 4. In 4 other latency windows of 3 experiments it was among the lowest probability group of models (the whole epoch analyses of experiments 4 and 5, and the 400 to 800 analyses of experiments 3 and 5).

Across the 4 category membership decision experiments, the 'Kucera-Francis 1' model is the least likely model to have been related to the ERP data by chance in the 400 to 800 msec analyses (see Figure 5.3). The only other models of comparably low probability are the 2 other word frequency models.

This model is among the worst predictors of ERP variance in the early and mid-latency analyses (see Figures 5.1 and 5.2). The statistically rare correspondence between this word frequency model and ERP data, therefore, seems confined to the later portion of the post-stimulus epoch. In this region of the epoch, though, there is a high probability that the linguistic factor captured by this model, word frequency, is an effective predictor of ERP variance.

So great is the variance associated with word frequency in the late part of the epoch that in 3 of the 4 decision experiments a word frequency model is the best predictor of ERP variance in the whole epoch analyses. In the fourth decision experiment (experiment 5) the word frequency models were among the best related to the whole epoch ERP data. This finding suggests that, in terms of the magnitude of the effects, the late word frequency effect is the largest and that it tends to occlude variance sensitive to other factors at different latencies.

Kucera-Francis 2: Like the word frequency model discussed above, this model was related to the ERP variance at a level unlikely to have come about by chance in several windows. It was associated with very low probabilities in 3 latency windows of 2 experiments (the whole epoch analyses of experiments 2 and 3, and the 400 to 800 analysis of experiment 2). It was among the lowest probability model group (the word frequency models) in a further 4 windows of 3 experiments (the whole epoch analysis of experiment 4, and the 400 to 800 msec analyses of experiments 3, 4 and 5).

Across experiments, this model was associated with the 3rd lowest probability for the 400 to 800 msec window, a probability only improved on by the fellow word frequency models, 'Kucera-Francis 1' and 'Kucera-Francis 3'. Hence, this model provides very similar evidence to that reviewed above in relation to 'Kucera-Francis 1', that word frequency is the salient factor late in the epoch.

Kucera-Francis 3: 'Kucera-Francis 3' was a close covariate of 'Kucera-Francis 1' in both sets of words. Its pattern of variance explanation was correspondingly similar. The model explains variance to a degree which is very unlikely to have occurred by chance in 3 windows of 2 experiments (the whole epoch analysis of experiment 2, and the 400 to 800 analyses of experiments 2 and 4). In 2 of these windows the model was associated with a probability of 0.01. In addition, the model was among the (word frequency) group best related to the ERP variance in 4 windows of 3 experiments (the whole epoch analyses of experiments 3 and 4, and the 400 to 800 msec analyses of experiments 3 and 5).

Across experiments, the model was a very poor descriptor of ERP variance in the early and mid-latency analyses. It was, however, a very good one for the 400 to 800 window, being second only to 'Kucera-Francis 1' (by only 0.01) in its improbability of having corresponded by chance. The model, as for the two preceding word frequency models, supports the conclusion that word frequency is the important factor late in the epoch.

Number of letters: This model was associated with a low probability in 4 latency windows (the 1-250 msec window of experiment 1, and the 250-400 windows of experiments 2,4 and 5). In addition, the model was a member of the group of models least likely to have been related to the ERP data by chance in 2 windows (the 1-250 msec analyses of experiments 2 and 4).

Across experiments the model is associated with about the same probability in the 1-250 msec and 250-400 msec analyses. In the early window, over all the experiments, it is by far the least likely model to have corresponded to the ERP data by chance. In the mid-latency analyses it was not the most successful model.

These features suggest that word length, the linguistic factor captured by 'number of letters', is the factor responsible for ERP differences in the earliest window. But they also suggest that the influence of word length continues beyond 250 msec, since the model fitted well in the succeeding 250-400 msec window. The model was not a close covariate of the most successful model in the mid-latency analyses ('semantic similarity', see Figures 3.7 and 4.4), so it is unlikely that its performance in the later window could be solely due to shared variance with the effective factor.

These considerations suggest strongly that word length is the factor which best predicts early ERP variance. The time course of the sensitivity of ERP variance, to this and other factors, is addressed below.

Number of syllables: 'Number of syllables' was associated with a low probability in only 1 window (the 1-250 msec window of experiment 4). It was a member of the group of models most unlikely to be related to the ERP data by chance in a further 2 windows (the 1-250 msec analyses of experiments 1 and 2).

Across experiments, the model is least likely to have been related to the ERP variance by chance in the 1-250 msec window, where it is second equal to 'Number of letters' as most successful model.

The performance of this model provides complementary, but less strong, evidence that word length is the effective factor early in the epoch. The lack of strong correspondence, for the most part, in the performance of this model may be related to the fact that the model possesses very little intrinsic variance for the words selected: All the words are either of 1 or 2 syllables. It may be that this model only crudely captured the linguistic factor of word length to which it was intended to relate.

Number of phonemes: This model yielded a very low probability in only 1 window (1-250 msec, experiment 2). However, it was a member of the most successful group of models in a further 2 windows (1-250 msec analyses, experiments 1 and 4). Across all the experiments 'number of phonemes' was about equally successful in the 1-250 and 250-400 msec analyses. In the former window across 5 experiments it was the equal second least likely model (with 'number of syllables') behind 'number of letters'. The 3 word length models are therefore the least likely to have been related to the ERP variance by chance in this window.

Phonological similarity: The lowest probability with which this model was associated in any analysis was 0.11, in the mid-latency analysis of experiment 5. Across all the experiments, its best performance was in the late time window. In this window it was below the average probability for all the models, but was only half as improbable by chance as the most successful, word frequency, models. This may have been due to a slight covariation with the word frequency models (see Figures 3.7 and 4.4).

There is no evidence from the results of the quantitative modelling experiments that this factor was important.

Orthographic similarity: Aspects of the performance of this model across the experiments have been discussed above. It was related to the ERP variance at a very chance-improbable level in only 1 window (the mid-latency analysis of experiment 5). In addition, it was related at a slightly less improbable level in a further 2 windows (250-400 msec experiment 4 and 1-250 msec experiment 3). As treated above, this pattern of performance is best explained by the fact that the model was a covariate of word length for the colour names, and a covariate of semantic similarity for the furniture terms.

Across experiments the orthographic model performed best in the mid-latency analysis, in which it was the third most successful model. It was associated with about twice the probability that it related to the ERP data by chance than the most successful model. The relatively good performance of the model in this latency window may be accounted for, again, by the fact that both its covariates in the two word sets are successful at explaining ERP variance in this window: 'Number of letters' and 'semantic similarity' are the 2 best fitting models in the mid-latency analysis.

No persuasive case can be made, on the basis of the data reviewed above, that orthographic similarity was an important factor. Indeed some aspects of the results (discussed under 'dependencies between the linguistic factor models' above) would not be readily explainable if this model had been an important factor.

Semantic similarity: 'Semantic similarity' was sufficiently precisely fitted to the ERP global variance model that it was associated with very low probabilities that the correspondence could have come about by chance in 5 latency windows of 4 experiments (the mid-latency analyses of experiments 2, 4 and 5, and the whole epoch and 250-900 msec analyses of experiment 1). Three of these analyses yielded probabilities of 0.01 (the whole epoch and 250-900 msec analyses of experiment 1, and the the 250-400 msec analysis of experiment 5), and the remaining 2 gave 0.02 and 0.04 (the mid-latency analyses of experiments 2 and 4, respectively). In only 1 experiment did it not fit the ERP data well in the expected latency window (the mid-latency analysis of experiment 3).

'Semantic similarity' was the only linguistic factor model ever to explain more than 50% of the variance in the global ERP model in any analysis. This it did in 4 windows of 3 experiments (the 250-400 msec analyses of experiments 2 and 5, and the whole epoch and 250-900 msec analysis of experiment 1). It was associated with the highest variance-explained value recorded in these experiments: 0.57 in the 250-900 msec analysis of experiment 1, a goodness of fit which is comparable to that for the relation of Cloze probability to N400 amplitude (Kutas and Hillyard 1984) since it is equivalent to a correlation of 0.75.

Across experiments, this model was very poorly fitted to the ERP data in the early and late time windows. In the mid-latency analyses, however, it was the best fitted model, being associated with about half the probability that its precision of fit could have come about by chance than its nearest competitor ('number of letters', see Figure 5.2), despite the likely presence of word length effects at this latency.

The probability that this pattern of results could have come about by chance is vanishingly small. The semantic similarity of the presented words, in both the colour and furniture categories, was the dominant linguistic factor underlying ERP variance in the mid-latency region of the post-stimulus epoch.

Conclusions

The results of the application of the quantitative modelling technique to ERPs evoked by exposure to single words suggest that 3 linguistic factors underlie differences between the ERPs. These linguistic factors are word length, semantic similarity and word frequency.

The relative strength of the evidence for the presence of each factor could be construed in a number of different ways. One way would be to take the disparity across experiments between the improbability of the most successful model, and the nearest model which is not related to the same factor, and to express this as a ratio. According to this criterion, the word frequency effect in the late window is the most strongly indicated ('Kucera-francis 1' is 2.2 times less likely to be related by chance to the ERP data than the nearest non-frequency model, 'phonological similarity'). This is followed by the word length effect in the early window ('number of letters' is 2.0 times less likely than 'mean letter frequency'). Least strongly indicated, according to this construal, is the semantic effect at mid-latencies ('semantic similarity' is 1.7 times less likely than 'number of letters') (see Figures 5.1 to 5.3).

Another construal might be to stress the number of times across the experiments that one model, associated with a particular linguistic factor, is accompanied by a low probability in the appropriate windows. If this is construed as the important feature, then the semantic model is the most strongly indicated as it its probability is below 0.04 in 5 appropriate windows. The word length and word frequency effects are equal second, as a model representative of each factor is at a probability of 0.04, or lower, in 3 windows.

No doubt other criteria could be applied. But whichever way the evidence is construed, it is very hard to account for the pattern of the results across the 5 experiments by explanations other than those implicating the action of these 3 factors. Explanations due to sequence effects fail on the fact that different sequences were employed but that these resulted in the same factors fitting well. Additionally it is hard to understand how the same sequence effect could give rise to different organizations, bearing a strong correspondence to those expected if these factors were operating, at different latencies. Explanations which rely on the action of pure chance endow the experimenter with an inordinate amount of (bad) luck. Explanations which suggest that the pattern of results is some artifact of the data reduction and analysis techniques must explain how it is that, when the source of the data changes (i.e. different latency windows), while the analytic techniques remain identical throughout, the results covary with these changes. The results cannot be imputed to the analytic techniques which were a 'control' factor. Explanations which rely on some alternative combination of linguistic factors must account for the problems associated with each alternative, set out above.

In addition to the identification of these 3 linguistic factors, the results also suggest that the maximal effect of each factor occurs at a different latency. The evidence is that word length influences the waveform first, followed by semantic similarity, which is then followed by word frequency. If the possiblity that the word length effect continues into the mid-latency window is allowed (see below), then it is the case that models related to these 3 linguistic factors *never* explained ERP variance well in inappropriate windows. For example, 'semantic similarity' never explained variance well in the early or late windows of the category membership decision experiments, nor did any word frequency model fit well in the early or mid-latency windows of the same experiments. This orderly feature of the results of these quantitative modelling experiments is another source of support for the suggestion that systematic relations between the linguistic and electrophysiological domains have been uncovered by these means. In order further to investigate this temporal aspect of the relations of these models to the ERPs, temporally more fine-grained analyses were required.

What is the time course of the sensitivity of ERP variance to the probable factors?

The results of some of the experiments, particularly experiments 2, 4 and 5, suggested that the influence of word length on the ERP waveforms might straddle the 250 msec partition. In order to examine these and other temporal aspects of the relation of the identified linguistic factors to the ERP waveform, a number of shorter latency windows were submitted to configurational analysis.

The width of the latency windows of these more fine-grained analyses was chosen to be 100 msec. This choice was made on the grounds that 100 msec was narrow enough usefully to localize an effect in time, but probably broad enough for the effects not to be overwhelmed by higher frequency noise (though noise obviously will be more of a problem the shorter the time windows). Analyses were conducted on 8 windows of this width beginning with a 50 to 150 msec window and ending with a 750 to 850 msec window. These latency windows may not optimally capture the time course of the effects, and the partitions may not optimally segregate variance due to each factor. But the shorter windows should give a clearer indication of both the time course and the relative influence of each factor on the waveform in these latency windows.

Full analyses on all the 13 models, for all 8 windows across 5 experiments, would require 520 randomization tests, or 208,520 PROCRUSTES rotations (for comparison, the analyses reported to this point required 103,458 rotations). Since these new analyses were intended only to focus on the relative time course of the sensitivity of ERP variance to each of the 3 identified factors, I decided to reduce this huge amount of computation to more manageable proportions. This was accomplished by dispensing with the randomization tests for these analyses. The relative goodness of fit of each candidate linguistic factor model is therefore reflected in the variance it explained in the ERP data, rather than in the rarity of its correspondence as determined by randomization. Each of the 13 linguistic factor models were tested against ERP data drawn from these latency windows. But, for clarity, the performance of the 'semantic similarity' model, the 3 'Kucera-Francis' models and 'number of letters' and 'number of phonemes' are reported. 'Number of syllables' was excluded from the set of word length models on the ground that it probably did not possess sufficient variance to usefully track variance related to word length (see above). The variance-explained values of the other models differed over the selected windows of the epoch, as would be expected from their covariance with the identified factors.

Tables 5.1 to 5.5 show the variance-explained values for the six models which represent the 3 identified linguistic factors for experiments 1 to 5, respectively. The R^2 values in these tables present a rather noisy picture, but one which seems consistent with the 'regional' analyses reported in Chapters 3 and 4, and summarized above. Interpretation of the results associated with

Table 5.1

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows of experiment 1 (6 subjects).

	Latency windows							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.280	0.201	0.417	0.071	0.270	0.245	0.365	0.396
'Kucera-Francis 1'	0.134	0.012	0.039	0.041	0.130	0.205	0.007	0.007
'Kucera-Francis 2'	0.001	0.003	0.027	0.033	0.176	0.281	0.132	0.204
'Kucera-Francis 3'	0.091	0.011	0.043	0.084	0.134	0.224	0.019	0.024
'Number of letters'	0.220	0.168	0.306	0.397	0.084	0.040	0.196	0.231
'Number of phonemes'	0.028	0.176	0.162	0.088	0.050	0.025	0.023	0.073

Table 5.2

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows of experiment 2 (6 subjects).

	Latency windows							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.041	0.100	0.273	0.385	0.111	0.348	0.375	0.109
'Kucera-Francis 1'	0.052	0.013	0.061	0.063	0.059	0.228	0.250	0.284
'Kucera-Francis 2'	0.069	0.038	0.100	0.030	0.042	0.350	0.292	0.227
'Kucera-Francis 3'	0.041	0.022	0.068	0.055	0.044	0.270	0.281	0.273
'Number of letters'	0.206	0.451	0.441	0.259	0.062	0.126	0.069	0.250
'Number of phonemes'	0.206	0.441	0.262	0.161	0.030	0.066	0.052	0.239

Table 5.3

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows of experiment 3 (6 subjects).

	Latency windows							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.297	0.338	0.259	0.170	0.244	0.370	0.181	0.509
'Kucera-Francis 1'	0.119	0.117	0.166	0.116	0.207	0.183	0.219	0.059
'Kucera-Francis 2'	0.003	0.129	0.120	0.148	0.173	0.257	0.189	0.164
'Kucera-Francis 3'	0.074	0.137	0.192	0.119	0.218	0.212	0.228	0.089
'Number of letters'	0.088	0.132	0.116	0.172	0.229	0.083	0.053	0.167
'Number of phonemes'	0.170	0.103	0.207	0.173	0.350	0.088	0.015	0.051

Table 5.4

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows of experiment 4 (7 subjects).

	Latency windows (msec)							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.114	0.237	0.322	0.553	0.198	0.054	0.098	0.146
'Kucera-Francis 1'	0.169	0.069	0.060	0.052	0.123	0.028	0.196	0.212
'Kucera-Francis 2'	0.134	0.067	0.079	0.071	0.045	0.176	0.193	0.244
'Kucera-Francis 3'	0.115	0.082	0.058	0.068	0.102	0.053	0.196	0.238
'Number of letters'	0.019	0.066	0.139	0.154	0.103	0.066	0.209	0.062
'Number of phonemes'	0.024	0.051	0.189	0.184	0.093	0.048	0.070	0.011

Table 5.5

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows of experiment 5 (8 subjects).

	Latency windows (msec)							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.198	0.241	0.227	0.247	0.311	0.116	0.168	0.238
'Kucera-Francis 1'	0.067	0.122	0.306	0.239	0.057	0.022	0.061	0.056
'Kucera-Francis 2'	0.011	0.131	0.212	0.169	0.033	0.072	0.052	0.050
'Kucera-Francis 3'	0.049	0.137	0.273	0.210	0.049	0.019	0.060	0.053
'Number of letters'	0.031	0.127	0.310	0.322	0.156	0.050	0.069	0.087
'Number of phonemes'	0.020	0.061	0.101	0.194	0.164	0.119	0.086	0.082

each experiment in isolation from the pattern across experiments suffers from all the problems identified above. Accordingly, Table 5.6 shows values of variance-explained statistics which are the average of the value for each model in each window of the 4 category membership decision experiments. Experiment 1, which employed a passive exposure paradigm, was not directly comparable with the succeeding experiments, and was not used in the derivation of the average variance-explained values.

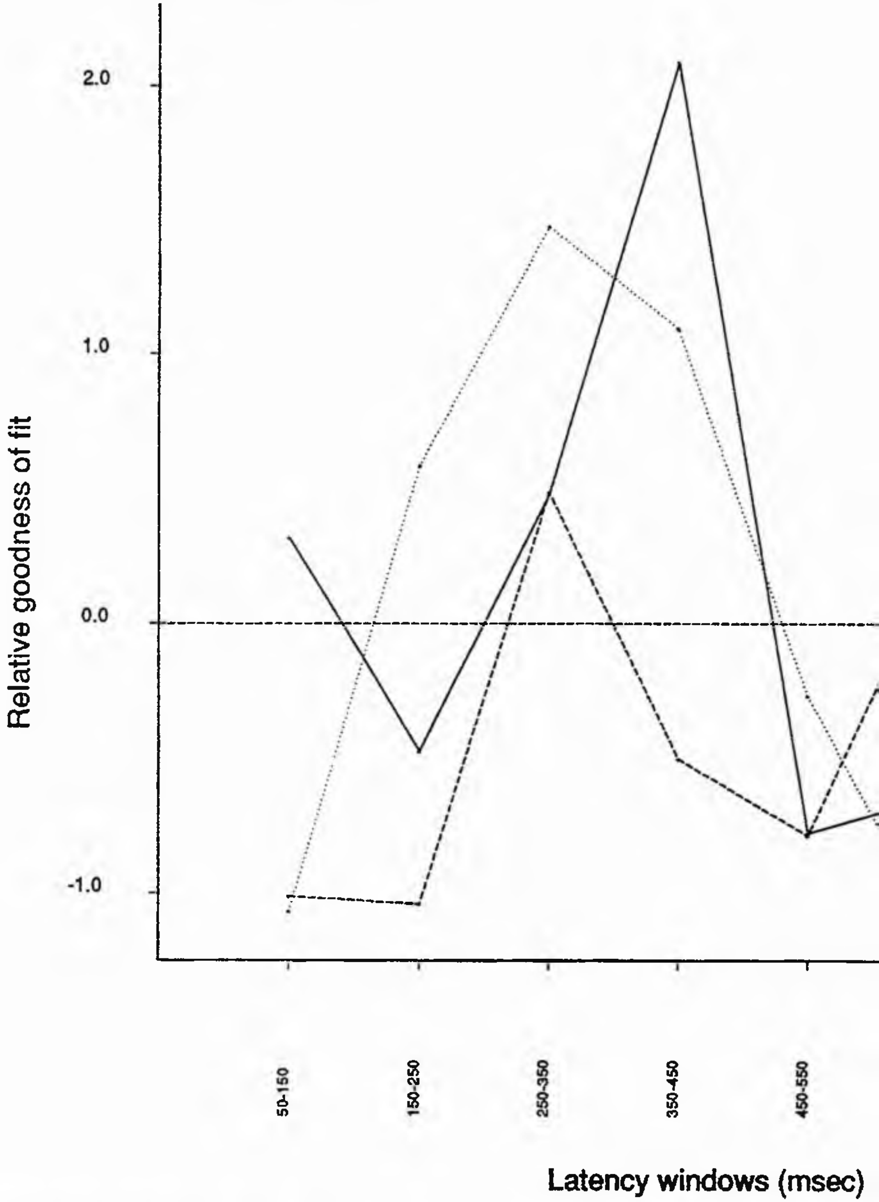
Figure 5.4 is intended to show clearly the time course of the 3 linguistic factors, as identified by these 100 msec analyses. It shows values taken from Table 5.6 for 'semantic similarity', 'Kucera-Francis 3' and 'number of letters'. These models were chosen to typify the pattern of variance explanation associated with each linguistic factor. The other word frequency models, and the other word length model, can be seen to closely covary across time windows with the chosen models (see Table 5.6). The following transform was applied to the values for the 3 chosen models. Because the 3-dimensional 'semantic similarity' model always tends to explain more variance than either of the unidimensional models, the mean of a model's values over all the time windows was subtracted from the value for that model in each window, to remove this advantage. The Figure therefore shows the relative capacity of

Table 5.6

Variance-explained values for models representing the 3 identified linguistic factors against the global ERP variance models for 8 successive 100 msec latency windows across experiments 2, 3, 4 and 5.

	Latency windows (msec)							
	50-150	150-250	250-350	350-450	450-550	550-650	650-750	750-850
'Semantic similarity'	0.263	0.229	0.270	0.339	0.216	0.222	0.205	0.251
'Kucera-Francis 1'	0.102	0.080	0.148	0.117	0.112	0.115	0.182	0.153
'Kucera-Francis 2'	0.079	0.091	0.128	0.105	0.073	0.214	0.182	0.171
'Kucera-Francis 3'	0.095	0.094	0.148	0.113	0.103	0.139	0.191	0.163
'Number of letters'	0.086	0.194	0.252	0.227	0.138	0.081	0.100	0.167
'Number of phonemes'	0.105	0.164	0.190	0.178	0.159	0.080	0.056	0.096

Time course of relation of the 3 identified linguistic factors to ERP variance across experiments 2 to 5.



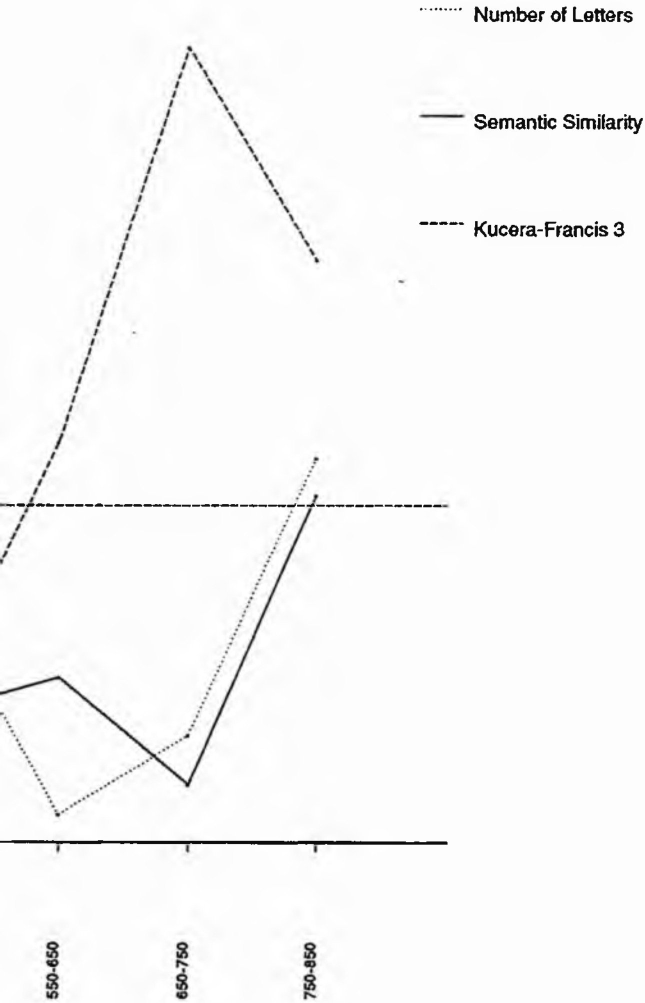

Number of Letters
Semantic Similarity
Kucera-Francis 3
550-650
650-750
750-850

FIGURE 5.4

each model to explain ERP variance across the latency windows. Because the values relate to a type of time series, they have been plotted in the form of a line graph.

The first feature of Figure 5.4 is that it shows that the performance, across experiments, of the 3 identified factors is peaked. The second feature is that the peak latency window of each model is different. The graph of the word length model rises more steeply than that either of the word frequency or semantic similarity models, and reaches a peak in the 250-350 msec window. The word length model explains little variance in the 50-150 msec window, substantially more variance in the 150-250 msec window and only slightly less variance in the 350-450 window than in the 250-350 msec analysis. The model explains variance relatively poorly in the later parts of the epoch. The effects of word length, according to these temporally more fine-grained analyses, were probably distributed across the 250 msec partition chosen for the 'regional' analyses presented in Chapters 3 and 4. Indeed, the analyses suggest that the word length effects began in the later portion of the 150-250 msec window, reached a maximum in the 250-350 msec window and were sufficiently prolonged in time to extend into the 350-450 msec window. The consequences of these results for the question of which features of the ERP waveform are likely to have mediated the word length effects are discussed below.

The semantic model explains variance best in the 350-450 msec window, rising sharply from the values for the preceding 250-350 msec window and declining sharply to the subsequent 450-550 msec window. The results suggest that variance sensitive to semantic similarity overlaps in time with variance sensitive to word length, in such a way that word length is the dominant factor in the window beginning at 250 msec and semantic similarity is the dominant factor in the window beginning at 350 msec. Hence it seems likely that the change in dominance of the two factors occurs somewhere in a latency window between 250 and 350 msec. The possible consequence of this feature of the results for the components of the ERP waveform likely to mediate the semantic effect is discussed below.

The word frequency model explains variance poorly in the first part of the epoch, but rises to a peak in the 650-750 msec window. This is consistent with the interpretation, drawn from examination of the waveforms of experiments 2 to 5 and from the 'regional' analyses reported in Chapters 3 and 4, that the word frequency effects are associated with the P600 deflection (see below).

What features of the ERP mediate variance related to the probable factors?

The word length effect

In the discussion of experiment 1 (see Chapter 3), I suggested that the word length model was more likely to be the effective model in the early analysis than the competing 'first letter frequency' model because word length might be influencing P200 through a 'size' effect (see Kutas and Hillyard 1983; Regan 1972). Although the evidence, drawn from all the experiments, that word length was the effective factor in that window of experiment 1 is much stronger than that for the letter frequency model, the explanation of the length effect as due to a modulation of P200 is problematic. P200 peaked at about 210 msec in all the experiments in Part One. The 100 msec analyses showed that the word length effect was maximal after that latency. Hence, the P200 is too early to be the component of the waveform solely responsible for mediating this effect.

It is possible that an N200 modulates the ERP waveform in a latency window beginning at, or just before, the P200 peak and continues for about 100 msec (see Rugg, Milner, Lines and Phalp 1987; van Petten and Kutas, in press). It would be possible to explain the apparent time course of the word length effect in these experiments if the same component were present here, and was sensitive to length. This component is reputed to be related more to aspects of visuo-perceptual processing than to linguistic processing per se (van Petten and Kutas, in press). Thus the word length effect, if mediated by an N200 component with the suggested time course, might still be a concomitant of a 'size' effect (see Chapter 6).

The semantic similarity effect

That there is a semantic factor implicated in ERP variance derived from experimental tasks which involve processing verbal material is not much in doubt (Bentin, McCarthy and Wood 1985; Kutas and Hillyard 1980a; 1980b; 1983; 1984; 1988; Kutas and van Petten, in press; Rugg 1987). The salience of word meaning, however, is typically indicated by the modulation of N400-like phenomena (see Chapters 1 and 7). These phenomena have been detected either in semantic priming paradigms (e.g. Bentin et al 1985; Rugg 1987), or in sentence processing paradigms in which the degree to which a particular word is expected on the basis of the prior context is manipulated (arguably also a kind of priming: see e.g. Kutas and Hillyard 1984).

The position in time of the semantic similarity effect, as indicated by the 100 msec analyses reported above, and its 'semantic' nature, are somewhat consistent with the presence of an N400, although the N400 is generally more extended in time than the apparent time course of the semantic similarity effect. A 'semantic' effect peaking at about 400 msec obviously suggests the possibility that an N400 was the component responsible for mediating the effect. But there are serious problems with the intuition that the semantic effect reported here is due to the presence of an N400: N400-like modulations due to semantic priming are probably abolished by repetition.

In a study which factorially crossed semantic priming with repetition, congruous-incongruous differences were diminished by one repetition and were completely absent in words which had been repeated twice (Besson, Kutas and van Petten 1989). The ERPs analysed above were derived, not from 3 presentations, but from 50 exposures (49 intra-experimental repetitions). Every colour word was, in consequence, thoroughly primed, both by repetition and by the context of the stimulus presentation, which was rich (obviously) in colour names. It seems unlikely, therefore, that the semantic effect in the present data could be imputed to a modulation of N400.

A second point militates against the interpretation that the semantic effect could be due to a modulation of N400. The modulation of this component indicates that it is affected by the degree to which an evoking word is semantically primed by the preceding context (Kutas and Hillyard 1984; see Chapter 7). If, for argument, I assume that semantic priming-related modulation makes a reappearance after a number of repetitions larger than that employed by Besson et al (1989) (which hardly seems likely), then to what extent would the colour words be *differentially* primed in this paradigm?

One possibility is that the semantic effect was mediated by an N400 sensitive to associative relations between word pairs in the experimental sequences (see Bentin et al 1985). This possibility can be excluded on the grounds, discussed above, that different sequences were employed across experiments, yet the experiments produced variance sensitive to the same linguistic factor. Similarly, there is a dissociation between the associative relations obtaining between items and the semantic similarity between them, as operationalized by semantic similarity judgements (Kutas and Hillyard 1988). An associative model would not have fitted well where the semantic similarity model fitted well.

Another possibility might be that, by being embedded in a presentation sequence of members of the colour name category, prototypical exemplars are primed to a greater extent than are atypical exemplars.

The first problem with this 'category priming' explanation is that all the chosen colour names and furniture terms were highly typical of the categories from which they were drawn (see the general introduction to Chapter 3 and the introduction to Chapter 4), so much so that (psychologically naive) subjects found it difficult to assign different values of typicality to the words (see Chapter 3).

The second problem is that this 'category priming' would probably result in a unidimensional pattern of variance in which prototypes are at one extreme, evoking a very small N400, and atypical items at the other, being associated with a larger N400. But this pattern of variance would not have been well fitted by the semantic similarity judgement models, which, in fact, fitted the ERP variance very well.

It seems improbable, on both empirical and conceptual grounds therefore, that the semantic factor derived here could be manifest in modulations of N400. There are no other candidate components which have been observed to intervene between an N250 and the P300 and which have a relation to the semantic properties of words. The component or feature of the waveforms responsible for mediating this effect is consequently unknown (see Chapter 8).

The word frequency effect

The late word frequency effect was associated with a time course consistent with the interpretation that it was mediated by the P600 deflection.

Conclusions and predictions for Part Two

The word length effect was the dominant effect in a latency window centred on a latency of about 250 msec. The predictions that can be made concerning this effect, on the basis of the results of these quantitative modelling studies, are as follows. There should be variance sensitive to word length in ERPs evoked by multiply repeated words. The main locus of these effects should be at a latency of about 250 msec. It is possible that the component which mediates the effect will be an N200-like modulation. The word length effects should be restricted to the earlier parts of the epoch since no relations between ERP variance and this linguistic factor were observed in the late portion. The possibility that word length-related variance is present in the late part of the epoch can only be consistent with the quantitative modelling studies if the variance due to this factor is smaller in magnitude than that due to the dominant factor (which should be word frequency). These predictions, derived from the results of my application of a new quantitative modelling technique to single word ERPs, are tested by a factorially designed experiment in Chapter 6.

The word frequency effect was the dominant effect in the late portion of the waveform. There should, in confirmatory experiments designed to test this aspect of the results, be variance sensitive to word frequency in ERPs derived from multiply repeated words. This variance should be confined to the later part of the epoch, particularly that associated with the P600 component, since variance sensitive to frequency was not found elsewhere. The possibility that variance sensitive to frequency is found at other latencies can only be consistent with the veracity of the results of the quantitative modelling experiments if its magnitude is less than that related to other factors (e.g. word length): Word frequency must not be the dominant factor in regions of the epoch other than the late region. These predictions, from the results of the quantitative modelling technique being applied to single word ERPs, are tested by a factorially designed experiment in Chapter 7.

The findings of a word length and a word frequency effect, in ERPs evoked by individual words, give rise to empirical predictions which are straightforward to test with conventionally designed and analysed experiments. As discussed in Chapter 1, corroboration of the veracity of the results of this new approach by analytic techniques which are the methodological norm is important. Specifically, such corroboration would both substantiate the empirical results themselves, and serve as an indication of the utility of the technique as a tool for empirical investigation in cognitive neuroscience. The finding of a semantic similarity effect, however, is much less straightforward to test by conventional means.

The veracity of the finding of the influence of any linguistic factor on the ERP waveforms can only be tested by conventional means if the factor segregates the linguistic domain into sufficiently large groups of words (see Chapter 1). This is for two reasons.

The first reason is that the conventional approach requires lists to be constructed which must be long enough to allow an ERP to be evoked. The ERP in one experimental condition is typically evoked by exposure to trials on each of which is presented one of the members of the relevant list. This necessitates lists of at least 40 items in practice.

The second reason is that control for extraneous factors must be effected by careful selection of the items which constitute the lists. That is, the members of each experimental list must be selected in such a way that they differ in the respect defined by the experimental factors, but do not differ in other respects. Hence, in order to test the effect of a particular linguistic factor on the ERP, in a properly controlled design, the factor must segregate the vocabulary into very large groups, so that a core of at least 40 items in each level of the factor remain when the items which differ in other respects have been discarded.

Both word frequency and word length are sufficiently 'global' factors that the partitions they effect in the vocabulary segregate huge groups of words. There are, for example, thousands of high and thousands of low frequency words. But semantic similarity judgements are not 'global' in the same sense. They are private to the category to whose members they relate, and are meaningless outside it. The factor captured by semantic similarity judgements does not segregate the vocabulary into anywhere near large enough groups of words from which to select controlled items. For example, the category of animal names is very large. Yet no systematic manipulation of the semantic similarity factor could be made which controlled for word length; the subset of animal names of the same number of letters is too small to derive one list of 40 items, let alone two lists so that a comparison could be made between levels of the semantic factor (see Chapter 6).

The very difficult problems associated with any attempt to investigate semantic similarity in a conventionally designed and analysed experiment have one positive aspect. They suggest that it has been possible to investigate a linguistic factor which is not explorable by a more conventional approach by means of the quantitative modelling approach developed for this purpose.

The conclusions drawn from experiments 1 to 5 concerning the semantic similarity effect may, however, be subject to a type of corroborative test by the conventionally designed experiments in Part Two. Corroboration of the effect of this factor will be dependent on the corroboration of the technique by the word length and word frequency experiments. The degree to which the evidence for the action of this factor is supported will be contingent on the degree to which the empirical results of the methodology applied in Part One are consistent with the empirical results of the methodology applied in Part Two.

PART TWO

Chapter Six

Word length and muliple repetition as determinants of the modulation of ERPs in a classification task

Introduction

Relation of the present experiment to the quantitative modelling studies in Part One:

This experiment was designed to test the conclusions, drawn in Part One, concerning the way in which word length modulates the ERP. Part One involved experiments in which an ERP waveform was evoked by each of a small set of words. Considerable quantitative information was generated about the words linguistic properties (Word length, letter frequency, phonology, orthography, semantic similarity etc.). The extent to which the structure of differences between the waveforms evoked by each word was in accord with the ordering expected on the basis of the words' linguistic dimensions was then assessed.

The waveforms were evoked by presenting each stimulus repeatedly (50 times) in a random presentation sequence. One finding was that, both when subjects were passively exposed, and when they were called upon to make a decision concerning the category membership of presented words, the modelling technique uncovered variance sensitive to word length in the pre-400 msec region of the epoch.

This length effect took the form of a region of length-sensitive variance extending from about 200 msec perhaps as far as 400 msec (see Chapter 5). By inference from the peak latency of P200, which was at about 210 msecs in all the experiments, the P200 could not have been the component of the waveform solely responsible for mediating the word length effect. Whatever the feature of the waveform which mediated this effect, the presence of a word length effect after the subjects had seen the stimuli very many times suggests that word length effects on the waveform were not abolished by repetition. Word length effects should therefore be present after repetition in the present experiment.

In order to fully validate the findings of the quantitative modelling studies two further features are required to be demonstrated. These features are that after repetition, length effects should not only possess a time course consistent with that discussed in Chapter 5, but also that length effects should not be the dominant factor in other regions of the waveform, such as the late portion. Failure to establish these expected features in the present study would reduce the credibility of the results of the quanititative modelling technique, while finding all of them would constitute corroboration.

Studies of word length and word repetition

The effects of word length have been investigated in a number of paradigms. Recognition thresholds to centrally presented words are not influenced by length (Seymour and May 1981), although the same measure of performance is affected by length when the stimuli are nonwords (Richards and Hempstead 1973). The effects of word length on naming latency are small (Bub 1982; Forster and Chambers 1973) except when the familiar horizontal format of the words is disturbed, in which case the effects of length are substantial (Seymour and May 1981; Young and Ellis 1985).

A major focus of recent work, in which word length has been manipulated, has been the interaction of this variable with the visual field in which the stimuli were presented (e.g. Bruyer and Janlin 1989; Bub 1982; Young and Ellis 1985). Performance (report accuracy in a naming task) related to stimuli presented in the right visual field is very similar to that for presentations at the centre of the field of vision (Young and Ellis 1985). Performance related to presentation to the left visual field, however, is length dependent for both words and nonwords, even when the words are presented in the usual horizontal format (Young and Ellis 1985).

Although the insensitivity of the normal reading system to word length has made length seem a rather uninteresting variable, the pattern of word length effects described above can inform theoretical accounts of word recognition (e.g. Young and Ellis 1985). Length effects are large for pronounceable nonwords, for words in unusual formats (e.g. $^{\mathrm{h}}{}_{\mathrm{o}}\mathrm{r}_{\mathrm{s}}{}^{\mathrm{e}}$) and for conventionally formatted words presented to the left visual field. Length effects are small for conventionally formatted words presented to the right visual field or to central vision. Young and Ellis (1985) organize these aspects of word length effects by supposing that there are 2 qualitatively distinct routes by which entries in an internal lexicon may be accessed.

There is, first, a 'whole word' route which is particularly characteristic of normal reading and of right visual field presentations. This route is not sensitive to word length. Second, there is a route which involves the participation of a short term graphemic representation of the stimuli (Young and Ellis 1985). This route is apparent in the event that the whole word route is unavailable (or delayed), as will be the case for nonwords, unconventionally formatted words and words presented to the left visual field (Young and Ellis 1985). Access to lexical entries is thus conjectured to be effected in a different way by verbal stimuli when their processing is mediated via the left or right hemisphere.

This account of word length effects reflects a general expectation that such effects will be present early in the processing sequence. The possibility, however, that there might occur some postlexical orthographic or letter level processing, perhaps a 'check' of the kind noted in Chapter 3, cannot be excluded.

Word length has been manipulated in a number of different ways. Young and Ellis (1985), for example, manipulated both 'visual length', operationalized as number of letters, and 'phonological length', operationalized as number of syllables. It was found that number of letters was much the better means of manipulating word length in a naming task in which performance was measured in terms of report accuracy. The efficacy of the number of letters was replicated by Bruyer and Janlin (1989), but with the additional finding that the number of letters was the important factor rather than the physical extent of the words. Blank spaces were inserted into 4 letter words so that their physical extent matched that of 7 letter words, with the result that the 'enlarged' 4 letter words were processed according to the number of their constituent letters rather than their physical extent (Bruyer and Janlin 1989).

With respect to the many studies of word repetition, the most typical finding is that repeating a word within the context of an experiment enhances the speed and accuracy of performance relative to unrepeated words (Scarborough et al 1977; Monsell 1985). To my knowledge, word length and repetition have not before been crossed in a factorially designed experiment. A much more detailed treatment of repetition effects is found in Chapter 7, where their possible relationship to word frequency effects more naturally places them.

In order to address the issues raised by the results of the quantitative modelling experiments, and to place the effects of word length on the ERP in the context of previous work, I decided on the following features for the design of this experiment. Words of 2 different lengths were chosen, these lengths being 4 letters and 7 letters. Four presentations of each of the words were included to assess the effects of initial repetition and subsequent multiple repetition on the length effects.

In order to be as directly comparable to the quantitative modelling studies in Part One, I wanted to employ a category membership decision task. In experiments 2 to 5 the category membership decision involved response being made to out of category targets (e.g. 'respond to any word which is not a furniture term'). There are, however, no categories large enough to provide a pool of words from which could be selected sufficient 4 and 7 letter items. This is particularly the case since control must be effected for (at least) word frequency. Hence I investigated the possibility of a paradigm in which the category membership decision involved response being made to 'in category' targets (i.e. 'respond to any word which is an animal name').

The target words in such a paradigm required to be either 4 or 7 letters in length, being the two levels of the word length manipulation in this experiment. The nontarget items would be of 4 or 7 letters; if the targets differed in length, the subjects could accomplish the task by employing cues related to the physical properties of the stimuli. But even the smaller number of items required to make up the target word set (80 items, see below) exhausts the capacity of any one category to provide 4 and 7 letter items. A category membership decision task therefore could not be employed here.

Instead I chose to employ a word-nonword (lexical) decision task. The set of 4 and 7 letter nonwords is very large, and it was possible to select sufficient target items from this set. A lexical decision task possesses a different task structure to that of the quantitative modelling experiments in Part One (see Chapter 7). But, at least with respect to the nature of ERP repetition effects, the ERP is modulated in a lexical decision task in a very similar way to that in a category membership decision task (Rugg, Furda and Lorist 1988). Also, if the finding of a word length effect were of insufficient generality to survive this change of task, I thought it unlikely that the findings of Part One would be of sufficient generality to bear on the intended issues. Nonetheless, the change of task from that of the quantitative modelling experiments remains a possible source of inconsistency.

Issues of experimental control in a conventionally designed and analysed 'across-stimulus' experiment

Where an experimental manipulation relates to an intrinsic property of the stimuli, such as word length, it is plainly not possible to counterbalance items in the experimental lists across conditions. This removes a major resource in controlling for extraneous sources of variance. In this situation, control for extraneous factors must be accomplished by careful item selection, a process which is made difficult by the fact that the determinants of variance in the ERPs from the individual words which constitute the lists are largely unknown. In Part One, however, I found that the major factors to which variance was sensitive in ERPs to repeated single words were word length, word frequency and semantic similarity. As an additional guide, Rugg (in press) controlled for initial letter and included only open class words.

Accordingly, it was decided to effect control in this study for initial letter, word frequency and for a general semantic factor, operationalised along the open-closed class distinction, and to assess possible item selection problems by including a second set of words as selection controls (see Bentin, McCarthy and Wood 1985).

Method

Subjects: Sixteen young adults were payed for participation. Seven subjects were female, all reported themselves to be right handed, to have learnt English as their first language and all were naive as to the purpose of the experiment. None was taking prescribed medication at the time of participation. Eight subjects had taken part in an evoked potential study before.

Stimuli: As discussed above, control was effected by item selection for word frequency and for a general semantic factor by including only open class words. Eighty words of 4 letters and 80 words of 7 letters were selected from the Kucera and Francis (1967) corpus. All items were selected in pairs. The pairs were of a short and a long word of similar frequency and with the same initial letter. All items were mid-frequency words of between 47 and 112 occurrences in the corpus. All the pairs save 2 began with the same initial letter. As far as possible, words which were derivatives of other word roots, such as participles and plurals, were excluded, but the pool of mid-frequency 4 and 7 letter words was not sufficiently large for these words to be avoided completely.

These 80 word pairs were further segregated into two sets. This was in order to employ one set as constituents of the main part of the design (which included repetition), and one set as a subsidiary part of the design to examine the adequacy of item selection (see below). This segregation resulted in two sets of 40 paired 4 and 7 letter words. The list of 4 letter words did not differ from the list of 7 letter words in word frequency for either set (paired t-tests: set 1 t =

0.76, p = 0.45; set 2 t = 0.54, p = 0.59). The word lists for this experiment are appended at Appendix 4.

Each member of the set of words which constituted the main part of the design was presented 4 times. Repeated items were inserted at lags taken from a rectangular probability distribution of median 10, and range between 6 and 14 intervening words, to minimise the predictability of where in the presentation sequence repetition might occur.

Each member of the set of words which constituted the part of the design related to control for item selection was inserted into the experimental list at random and did not repeat.

The target nonwords were derived from mid-frequency words of 4 and 7 letters which were not in the above sets. This derivation was accomplished by changing 1 or 2 letters of the words to yield pronounceable nonwords. Pseudohomophones were not allowed. The nonword targets were not repeated (to avoid the task becoming easier as the experiment progressed), and were inserted at random into the lists with a constant probability density (p = 0.17).

Four experimental lists were constructed in this way. Two lists were constructed with the role of the two sets of words reversed. For half the subjects the members of one set were the control items, and the members of the other set were repeated, while for half the subjects these roles were reversed. In this way the different item selections were counterbalanced across subjects. Two further lists were derived from these lists. In these new lists the serial positions of long and short words were reversed. This was in order to control for order of presentation or sequence effects. These lists were again counterbalanced across subjects.

Design and task: Subjects were asked to make a button press response with their preferred hand as quickly and accurately as possible when they detected a nonword. The nontarget words consisted of two levels of length (4 letters (S) and 7 letters (L)) crossed with four levels of repetition (first presentation and 3 repetitions (R0, R1, R2 and R3)).

Control for item selection was achieved by a subsidiary part of the design which crossed two levels of word length with two levels of word type, the one being the first presentations of words which would repeat and the other being the control items. As noted above, the identity of the two word types were counterbalanced across subjects. The two lists in which the order of presentation of the long and short words were reversed were counterbalanced across subjects. The design was thus completely counterbalanced with respect to order of presentation and word type.

The design is set out in Figure 6.1.

Features of the design and analysis of experiment 6

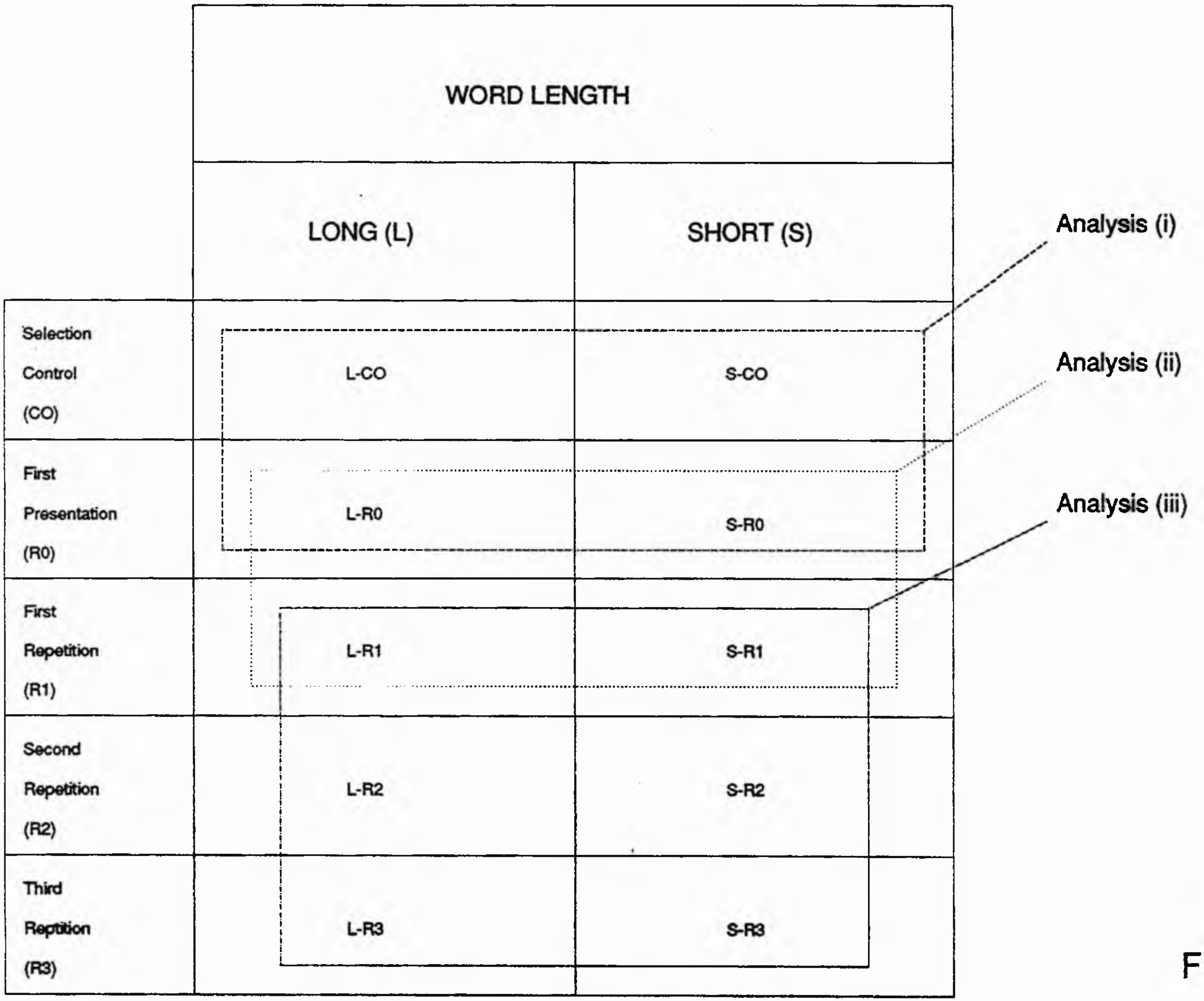


FIGURE 6.1

ERP recording:

All aspects of the electrode preparation and placement, amplification and digitization, stimulus presentation and averaging were identical to those detailed in Chapter 4.

Procedure: After application and impedance testing of the electrodes, subjects were seated in a soft upright chair facing a table on which were fixed two response buttons at right and left, and a TV monitor. The task was explained to be that of responding as quickly and accurately as possible to any stimuli which were not words.

It was explained that the stimuli would be presented on the screen one at a time but that the target nonwords would be mixed with non-target words as distractors. It was explained that some of the non-target words would repeat, but that this was irrelevant to the task of responding to the nonwords. The reason for the repetitions was explained to be laziness on the experimenter's part: Repeating non-targets made it possible to choose fewer words when designing the experiment. Subjects were also asked to try to withold blinking during each trial.

A sequence of 25 practice trials was administered, to ensure that the subjects were comfortable, understood the task and had no problem seeing the words, after which subjects had the option of altering their position or the lighting level. The practice trials were made up of presentations of a list derived from the target words from experiments 4 and 5, in which some repeats and nonwords were inserted. None of the words or nonwords in the practice sequence were repeated in the experiment.

When the subject was ready the experiment proper began, which was composed of 4 blocks of 120 trials, between each of which there was a short break. Subjects were debriefed during removal of the electrodes.

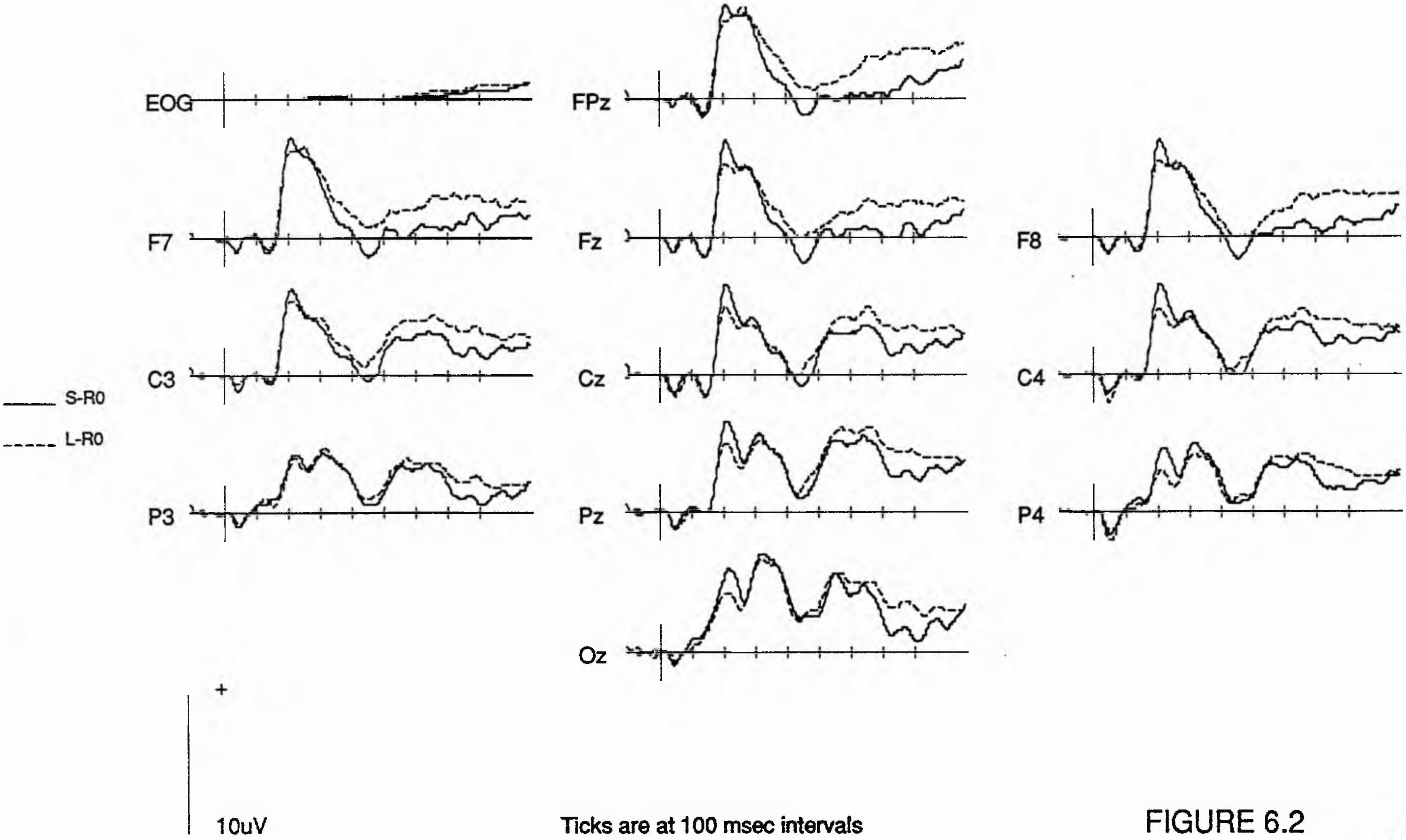
Results

Behavioural data: Mean Reaction time to the nonword targets was 690 msec, with a standard deviation across subjects of 92 msec. There was no significant difference between reaction times to the 4 and 7 letter nonwords (t-test $t = 0.19$, $p > 0.85$). The mean error rate for misses (errors of omission) was 2.4%, and the mean false positive rate was 0.3%. These error rates were considered too low to submit to further analysis.

ERP Waveforms

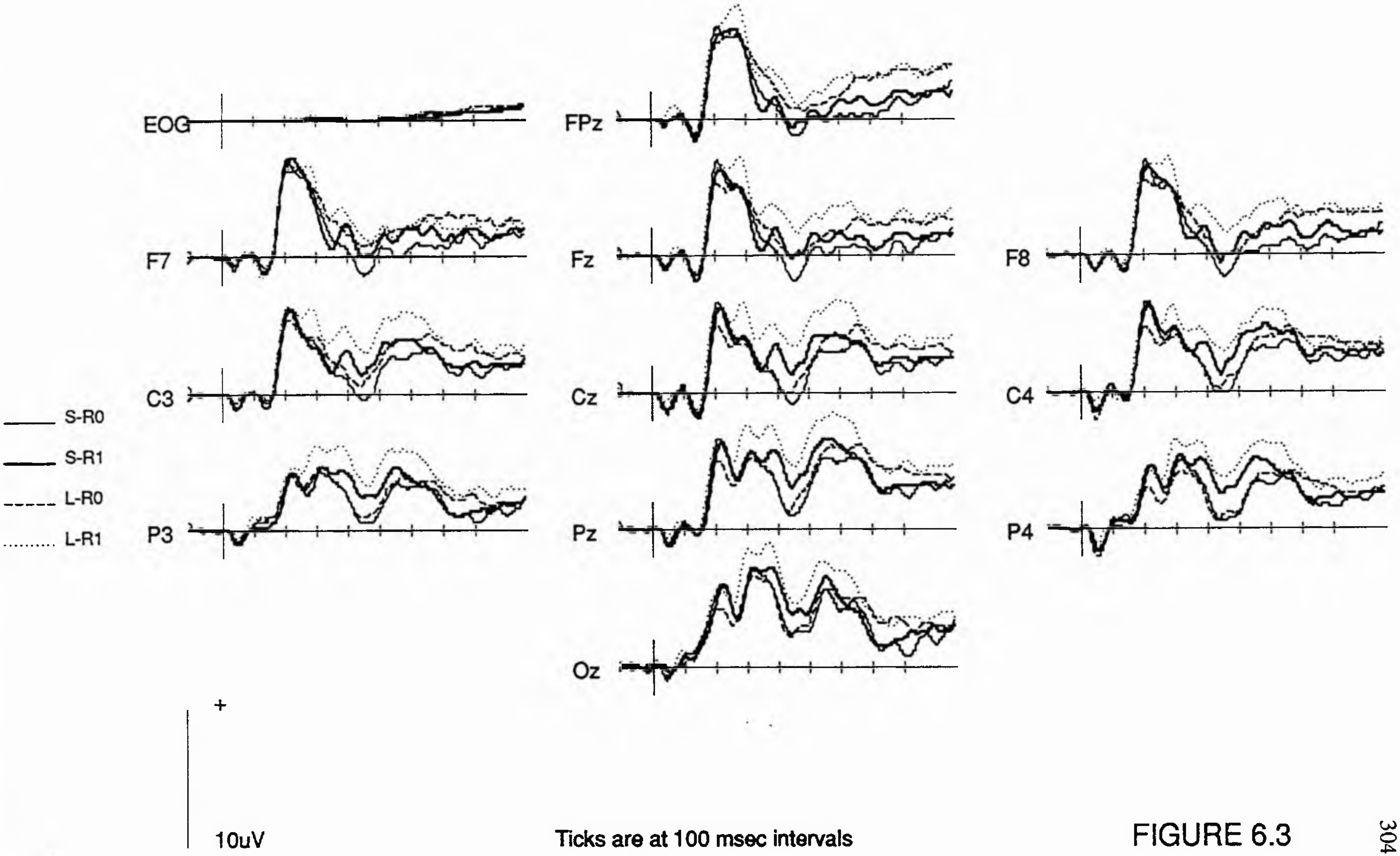
Grand average ERP's are shown in figures 6.2 to 6.5.

Figure 6.2 shows the waveforms evoked by the first presentations of both short and long words. The trace relating to the first presentation of the long words (L-R0) is deflected relatively more toward the positive in a latency window which overlaps the peak of the P200 component. This relative positivity in the long



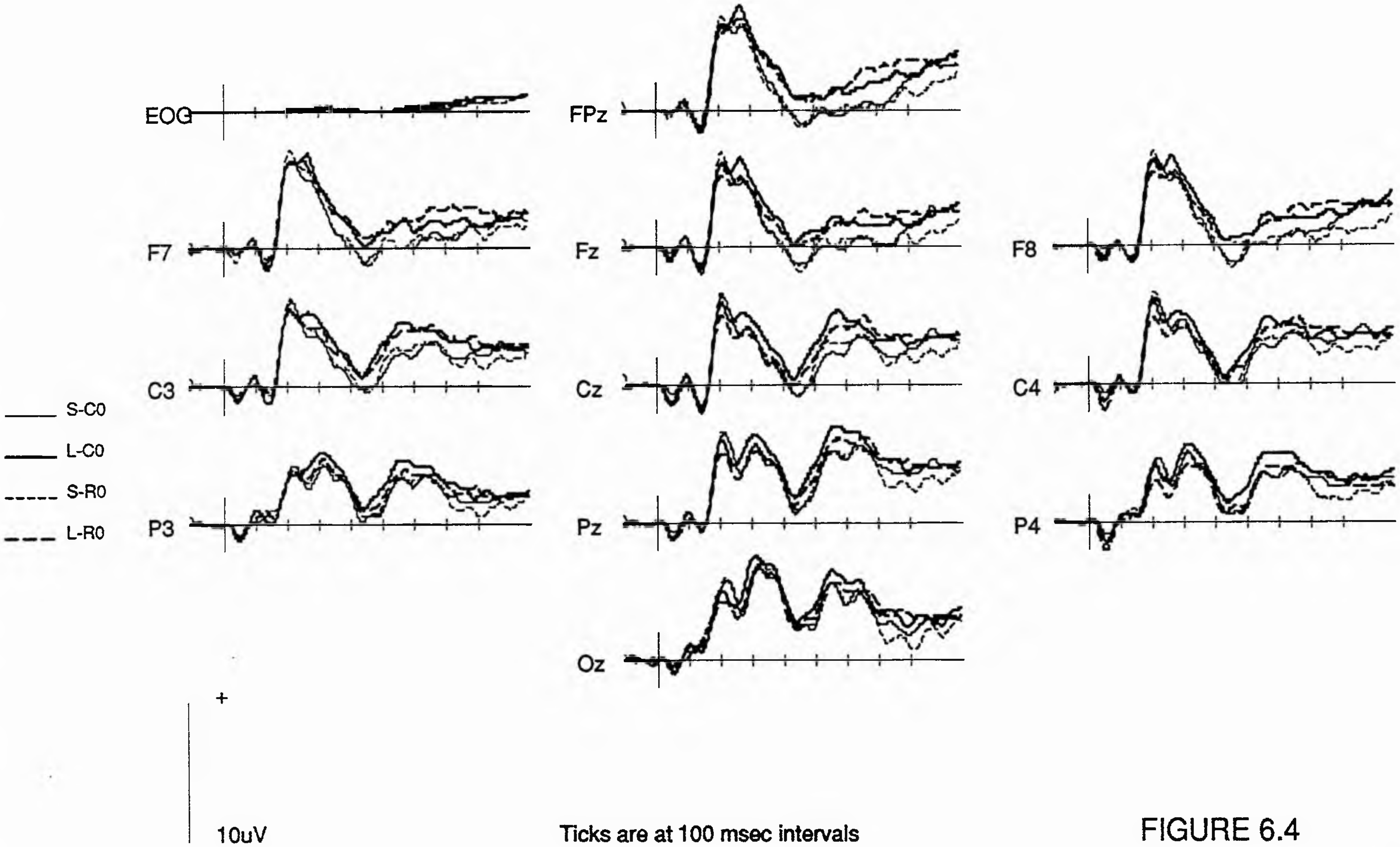
Ticks are at 100 msec intervals

FIGURE 6.2



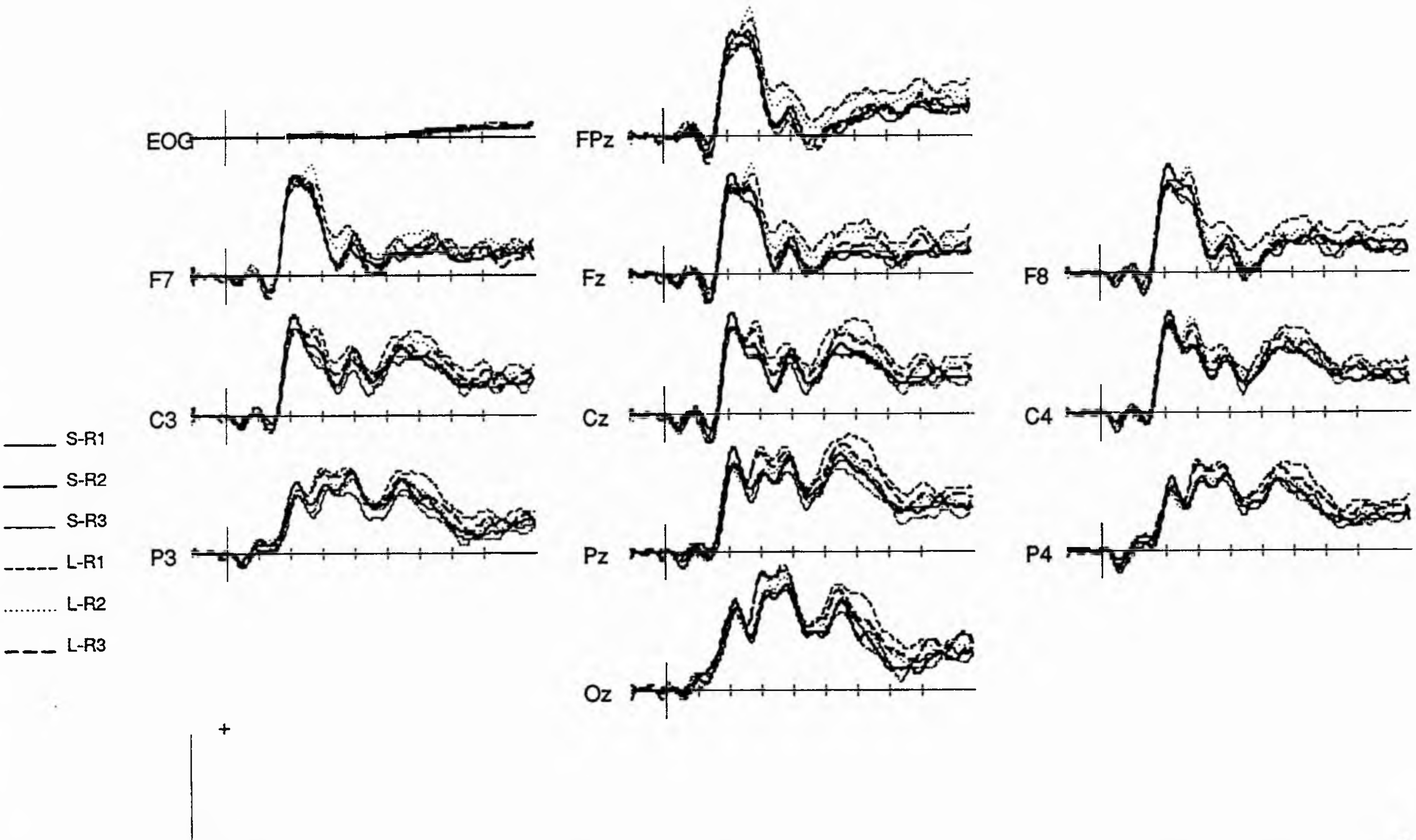
Ticks are at 100 msec intervals

FIGURE 6.3



Ticks are at 100 msec intervals

FIGURE 6.4



Ticks are at 100 msec intervals

FIGURE 6.5

words' first presentation trace is also apparent later in the epoch as a sustained difference, particularly noticeable on the frontal leads.

Figure 6.3 shows the waveforms evoked by the first presentation of both short and long words, together with the waveforms evoked by their first repetitions. Repeated words are characterised by greater positivity than is apparent in the first presentation waveforms, as has been demonstrated previously (Rugg 1985; Rugg et al 1988).

The short words' repetition condition (S-R1) begins to diverge from the S-R0 trace at about 350 msec and this relative positivity is sustained until just after 500 msec. The long words' repetition waveform (L-R1), however, begins to diverge from the L-R0 trace rather earlier, at about 200 msec, and has a more complex morphology. A large relative positivity is present which peaks at about 270 msec. This deflection declines at the frontal leads so that at about 300 msec the amplitude of the L-R1 ERP is similar to that evoked by the S-R1 condition. Centro-parietally, however, the relative positivity of the waveform of the repeated long words is sustained. The waveform is characterised in the latter part of the epoch by a large positivity, which peaks at about 600 msec.

Figure 6.4 shows waveforms evoked by the S-R0 and L-R0 conditions and the high and low frequency item selection control conditions (S-CO and L-CO, respectively). There is little difference between the traces on the basis of word type. The control waveforms can be seen to reflect the same somewhat frontally distributed length-related positivity mentioned in relation to the first presentation waveforms, S-R0 and L-R0.

Figure 6.5 shows the waveforms evoked by the three repetitions of the short and long words (S-R1, 2 and 3; L-R1, 2 and 3). There are visible differences between the 3 long word traces and the 3 short word traces. These are manifest in the fact that the long word traces are all more positive going than waveforms derived from repetitions of the short words at about 250 msec. This difference is particularly clear over the centro-parietal and left hemisphere leads. It is also possible that there is a length-related difference centred on the P600 deflection, though this is less clear than the early difference in this Figure.

ERP analyses

The data were measured by taking the mean amplitude of the waveforms relative to the 100 msec pre-stimulus baseline in selected latency regions. Time windows of 180-230, 230-360, 360-500 and 500-800 msec were chosen both to capture the apparent differences in the waveforms and to preserve some compatibility with other studies (e.g. Rugg, in press). Measurements in these time bands were submitted to repeated measures analysis of variance (ANOVA), with Greenhouse-Geisser correction of degrees of freedom (Keselman and Rogan 1980).

Analyses in each latency region were undertaken separately for midline and lateral sites to explicitly examine the possiblity of hemispheric asymmetries in the distribution of potential. In the midline analyses, 3-way ANOVAS were performed for 3 sets of factors. These analyses are set out in relation to the design in Figure 6.1.

Analysis (i) included factors of word type (control vs first presentations), word length and electrode site, in order to assess the adequacy of item selection control and to examine the effect of the word length manipulation on both control and R0 conditions.

Analysis (ii) employed factors of repetition (R0, R1), length (L, S) and electrode site. This analysis was undertaken in order to examine the effects of the length and repetition manipulations on initial repetition, and to be directly comparable with the many other studies of repetition effects (e.g. Rugg 1987; in press).

Lastly, analysis (iii) used factors of repetition (R1, R2 and R3), length (L, S) and electrode site. This analysis was undertaken to determine the effects of multiple repetition, and the degree to which length effects were present in repeated traces.

In the lateral site analyses, 4-way ANOVAS were conducted with the same sets of factors as for the midline analyses, but with the additional factor of hemisphere.

The mean amplitudes of each cell of the design (N = 16) are presented for the 4 latency bands which were subjected to analysis, in Table 6.1.

180 - 230 msec

The analysis (i) ANOVAs (electrode site x word type x word length) produced no effects related to the experimental manipulations. The analysis (ii) ANOVAs (site x length x repetition (R0 and R1), however, gave an interaction of length by repetition (F1, 15 = 12.7, p = 0.003) indicative of an increase in the positivity of the long word trace in the region of the P200, on repetition.

The analysis (iii) ANOVA gave no effects related to the experimental manipulations.

230 - 360 msec

The analysis (i) ANOVA for this window, which focussed on the unrepeated control and R0 conditions, indicated a length by hemisphere interaction (F1, 15 = 12.5, p = 0.003). This was related to greater length-related differences over the left scalp sites in this latency window. There were no effects related to word type.

Analysis (ii) ANOVAs, focussing on the first presentation of the words and their initial repetition, revealed main effects of repetition (F1, 15 = 5.22, p = 0.04) and of length (F1, 15 = 4.82, p = 0.04). In addition, there were a number of

Table 6.1

	4-Letter					7-Letter				
	Fz	Cz	Pz	C3	C4	Fz	Cz	Pz	C3	C4
180-240										
Control:	5.6	5.7	5.9	5.5	5.6	5.6	5.3	5.5	5.1	5.3
R0:	6.0	5.8	6.0	5.7	5.9	4.9	4.4	4.7	5.2	4.5
R1:	5.5	5.4	5.5	5.4	5.7	6.1	5.9	6.0	6.0	5.7
R2:	6.5	6.4	6.8	6.5	6.3	6.1	5.6	5.6	5.9	5.6
R3:	5.8	5.6	5.8	5.7	5.8	5.6	5.6	5.9	5.6	5.7
230-360										
Control:	3.7	3.3	3.6	3.3	4.0	5.0	4.5	4.6	4.5	4.6
R0:	4.0	3.5	3.9	3.5	4.0	4.7	3.9	4.0	3.9	3.7
R1:	3.4	3.5	5.0	3.4	4.2	5.3	5.3	6.7	5.2	5.3
R2:	3.8	3.9	5.5	4.1	4.1	5.1	4.8	6.2	4.9	5.2
R3:	3.4	3.6	5.3	3.7	4.0	5.1	4.7	6.5	4.7	4.8
300-500										
Control:	-0.6	0.2	2.3	0.4	0.7	1.3	1.9	3.5	2.1	1.8
R0:	-0.5	0.5	2.5	0.7	1.0	0.8	1.2	2.9	1.7	1.3
R1:	0.9	2.8	4.7	2.6	3.0	2.9	4.5	6.2	4.3	4.0
R2:	1.3	3.4	5.4	3.7	3.1	2.2	4.0	5.7	4.0	3.7
R3:	1.1	3.2	5.1	3.3	3.0	1.3	3.7	5.8	3.6	3.3
500-800										
Control:	0.5	2.7	4.6	2.5	3.2	1.9	4.0	5.6	3.8	4.0
R0:	0.6	2.7	4.2	2.5	3.1	2.3	4.0	5.4	3.8	4.1
R1:	1.4	3.4	4.8	3.2	4.0	3.2	5.2	6.5	4.7	4.8
R2:	1.5	3.6	4.8	3.5	3.6	2.5	4.3	5.1	4.0	4.3
R3:	1.7	3.3	4.1	3.1	3.5	2.2	4.4	5.7	4.0	4.3

interactions. The first was that of length by repetition (F1, 15 = 7.46, p = 0.02) which was associated with a greater increase in positivity on repetition for the 7 letter words. The second was a length by hemisphere interaction (F1, 15 = 4.73, p < 0.05), again indicative of greater length-related differences over the left scalp sites. The third interaction was a repetition by site interaction (F1.1, 17.2 = 5.51, p = 0.03), which showed that repetition effects tended to be greater over the parietal electrodes.

The analysis (iii) ANOVA, designed to test the effect of multiple repetition on the repeated traces, which used only the 3 repeated conditions (R1, R2 and R3), gave an effect of length (F1, 15 = 15.35, p = 0.001) but no effect of repetition nor any interaction of length with repetition (Fs < 1 in both cases). Hence the length effects in the repeated traces did not differ reliably as a consequence of multiple repetition in this latency window.

There was, however, an interaction of repetition with site (F3.2, 47.6 = 5.37, p = 0.002) indicative of the fact that repetition-related differences were present only at the frontal sites. In addition, there was a 3-way interaction of length with hemisphere and site (F1.8, 27.1 = 4.06, p = 0.03), related to the fact that length effects were larger over the frontal right scalp, but were equipotential over the left scalp.

360 - 500 msec

Analysis (i) ANOVAs, intended to examine length and word type effects in the first presentation traces, revealed a main effect of length (F1, 15 = 6.30, p = 0.02) and an interaction of length with electrode site (F1.2, 17.9 = 6.38, p = 0.02). This interaction pointed up the frontal distribution of the first presentation length effect at this latency (see Table 6.1 and Figure 6.3). There were no word type effects.

The analysis (ii) ANOVA gave main effects of length (F1, 15 = 7.00, p = 0.02), and of repetition (F1, 15 = 38.59, p < 0.001). There was also an interaction of repetition by site (F1.3, 19.4 = 15.28, p < 0.001), related to the greater repetition effect over the centro-parietal sites. For the lateral sites, there was a 3-way interaction between length, hemisphere and site (F1.8, 27.1 = 5.65, p = 0.01), due to the length effects tending to be larger over the central scalp for the left hemisphere, but larger over the frontal scalp for the right hemisphere electrode sites. There was no length by repetition interaction (F < 1.1) in this latency window.

The analysis (iii) ANOVA, based on the repeated conditions only, gave 2 main effects. These were that of length (F1, 15 = 8.34, p = 0.01) and that of electrode site (F1.6, 24.3 = 34.68, p < 0.001). Additionally, there were interactions between repetition and site (F3.8, 52.5 = 4.45, p = 0.005), due to repetition effects being present at the frontal leads. There was also a 3-way interaction between length, hemisphere and site (F2, 29.7 = 9.92, p = 0.001), related to the fact that length effects were larger over the left centro-parietal scalp, but larger over the frontal right scalp. Finally there was a 4-way interaction of length by repetition by hemisphere by site (F3, 44.6 = 3.50, p = 0.02), due to the diminution of length effects by multiple repetition being larger over the frontal left scalp than over the analogous right hemisphere sites.

500 - 800 msec

The ANOVA which examined word type and word length effects on the first presentation traces (analysis (i)) revealed a main effect of length (F1, 15 = 7.67, p = 0.01). This was indicative of the greater positivity in the long words' first presentation trace. There was no effect of word type.

Analysis (ii), examining the first presentation and initial repetition of the long and short words showed four main effects. There was a main effect of length (F1, 15 = 8.86, p = 0.01) reflecting greater positivity in the long word traces. There was a main effect of repetition (F1, 15 = 5.53, p = 0.03), due to the greater positivity of the repeated traces compared to the unrepeated traces. There was a main effect of hemisphere (F1, 15 = 7.42, p = 0.02), related to the fact that the right hemisphere sites were more positive than those on the left. Lastly, there was an effect of electrode site (F1.6, 23.9 = 11.31, p = 0.001), in virtue of the presence of a positivity with a parietal maximum. In addition, there was a 3-way interaction between length, hemisphere and site (F1.5, 23.2 = 6.01, p = 0.01) indicative of the fact that length effects over the left scalp were largest at the central and parietal leads, while over the right scalp length effects were larger at the frontal lead. There was no length by site interaction ($F < 1$), nor any interaction of length with repetition ($F < 1$).

An analysis based on the 3 levels of repetition (analysis (iii)) yielded 3 main effects. The first was that of length ($F1, 15 = 8.06$, $p = 0.01$) due to greater positivity in the 7 letter word waveforms. The second was an effect of hemisphere ($F1, 15 = 6.99$, $p = 0.02$) reflecting greater positivity over the right hemiscalp. The third was a main effect of site ($F1.5, 21.9 = 9.81$, $p = 0.002$) due to the parietally maximum positivity. There was no main effect of repetition in this latency window, nor an interaction of length with repetition.

This analysis also showed a 3-way interaction of length by hemisphere by site ($F1.9, 28.2 = 5.52$, $p = 0.01$) indicative of larger length effects over the central left scalp as opposed to larger length effects over the frontal right scalp. There was, in addition, a 4-way interaction between length, repetition, hemisphere and site ($2.5, 37.8 = 4.19$, $p = 0.02$) indicative of effects of very similar form to those signalled by the same 4-way interaction in the 360-500 msec analyses. There was no length by site interaction ($F < 1$).

Discussion

The results of this experiment were complex. My discussion of them follows a rough chronological order and is intended to begin with the simpler issues.

180 - 230 msec

This latency window was intended to bracket the ERP differences in the region of the P200 peak. The small differences which appeared to be present between the short and long word traces on first presentation (see Figure 6.2) were not reliable.

The interaction between length and repetition in this window was indicative of an increase in the positivity of the long word trace on repetition, while there was little change in the amplitude of the waveform derived from short word presentations (see Figure 6.3). This suggests two things. The first is that this feature of the results implies that word length did, indeed, modulate the ERP waveform. The second is that repetition and word length may share a processing stage which influences this earliest latency window.

The absence of any effects related to the experimental manipulations in the repeats only analysis, however, indicated that multiple repetition did not modulate the ERPs. Another implication of this result was that there were no length-related differences evident in the repeated traces at this latency. It seems unlikely, then, that the length-related variance detected in the early analyses of the experiments in Part One could have come from ERP differences in this window.

230 - 360 msec

The word length manipulation modulated the ERPs both before and after repetition in this window.

The interaction of length and hemisphere for the unrepeated conditions, characterised by greater length differences over the left scalp, implies both that length was a factor in the unrepeated traces and that there may have been differential engagement of the two cerebral hemispheres. This latter feature could accord with the supposition of Young and Ellis (1985) that there are different processing routes for verbal material depending on the hemisphere mediating the transmission of information. This asymmetry might imply that processors associated with both routes are engaged by centrally presented words, though only the faster, non-length-sensitive, route is apparent in behaviour, and that these processing differences are the cause of the asymmetric distribution of potential. The fact that length differences are greater over the left scalp, and not the right hemiscalp which overlies the hemisphere to which left visual field stimuli are projected, need not trouble this interpretation. Inferences about the location of the intracerebral generators of asymmetric ERPs are notoriously complicated (see Chapter 1) and lateralized generators can give rise to distributions superficially indicative of generators located in the opposite hemisphere (Arthur et al 1989; Barrett et al 1976). The presence of this length-related asymmetry can only be taken to indicate that the two cerebral hemispheres were engaged differentially by the length manipulation at a latency relatively soon after stimulus presentation.

The analysis which focussed on the first presentation and first repetition conditions showed a main effect of repetition, which, being characterised by an increase in positivity in the repeated traces, was consistent with the large number of previous studies which have investigated this variable (e.g. Bentin and McCarthy 1989; Nagy and Rugg 1989; Rugg 1985; 1987; in press; Rugg and Nagy 1989).

The absence of repetition effects in the 'repeats only' analysis implies that the effects of repetition were evident for the first repetition but that further repetitions did not modulate the ERP waveform. This finding is rather consistent with the view that repetition effects act by abolishing a 'default' N400 which is present in the waveforms to initial presentations (Rugg et al 1988) (but see Chapter 7 for a fuller discussion of this issue).

A repetition by length interaction was present in the analysis which compared first presentation and repetition for this latency window. The simplest interpretation of this interaction would be to suppose that repetition and length both act on the same processor. But the length manipulation interacted with hemisphere and the repetition manipulation did not. Repetition interacted with electrode site while length did not. Hence an interpretation of this pattern of interactions is to suggest, contrary to the simplest account, above, that these two factors modulated at least partially independent generators.

Despite the length by repetition interaction, when comparing first presentation and first repetition ERPs, the analysis which was based only on the repeated conditions produced a highly reliable effect of length. There was no effect of repetition, nor any interaction of length with repetition in this analysis. Repeated words which differ in length therefore gave rise to reliably different waveforms in the 230-360 msec region of the epoch. Length effects were present in repeated traces at the latency inferred in Chapter 5. Word length modulated the ERP waveform in just the way that was conjectured from the results of the experiments in Part One in this latency window.

360 - 500 msec

The main effect of length and the interaction of length with site in the analysis which examined the first presentation traces showed clearly the somewhat frontal distribution of the mid-latency length effects (see Figures 6.2 and 6.4). In contrast to the comparable analysis in the previous latency window there was no length by hemisphere interaction. This feature may be taken to suggest a dissociation between the length effects in the previous window and the present one on scalp distribution grounds.

The analysis which examined the first presentation and first repetition conditions showed a highly reliable effect of repetition as has been reported before (e.g. Rugg 1985). There was also an effect of length, but in contrast to the 2 earlier windows there was no interaction between length and repetition. The effects of these 2 variables were therefore additive in this mid-latency region of the epoch. This is apparent in Figure 6.3 which shows that the same length-related amplitude difference as is present in the first presentation traces is present in the repeated traces but deflected toward the positive by the presence of a repetition effect. The fact that these 2 factors were additive in this window and were interactive in the 2 early windows again implies a dissociation between the character of the experimental manipulations in this mid-latency window and their character in the early part of the epoch.

The presence of a first presentation length effect and the complete absence of length by repetition interaction in this analysis and the 'repeats only' analysis had the consequence that there were length effects present in the repeated traces at this latency. The only suggestion that the length effects in this latency range might be less robust in the face of multiple repetition than those in the 230-360 msec window was the high order interaction between length, repetition, hemisphere and site. This interaction is difficult to interpret, but may have related to the diminution of the length differences being larger at the frontal sites over the left hemisphere, but larger at the parietal sites over the right hemisphere.

500 - 800 msec

The first presentation conditions differed as a consequence of length in this latency window, as they did in the previous one. The effects of length were due to the sustained relative positivity associated with long words which was apparent in the analyses of the previous window. In this late region, however, the length effect was evidently more equipotentially distributed, since length did not interact with site in this window.

Analyses which focussed on the first presentation and initial repetition conditions gave main effects of length and repetition, but as in the 360-500 msec window there was no interaction between length and repetition. The same length-related positivity is present after repetition as before it.

The main effect of site in this window reflected the presence of a large positivity with a parietal maximum. It may be that this feature, which bears a strong resemblance to a visual P300 (P3b e.g. McCarthy and Donchin 1981), is only manifest in the waveform when a decision is required of the subject, since the first quantitative modelling experiment, using a passive exposure paradigm (experiment 1, Chapter 3), produced no such late positivity.

Before analysing the data in the way reported above, my intuition from examination of the waveforms was that the late length-related differences were mediated by this parietally maximal positivity, in the same way that the word frequency effects in Part One were mediated by a very similar deflection. But this is most unlikely for 2 reasons.

First, the length factor did not interact with site in any ANOVA for this latency window. Hence the length differences were not distributed as would be expected if the 'P300' were mediating the effect. The length effects are not reliably larger where the positivity is reliably larger (cf., Chapter 7).

Second, the actual distribution of length effects, revealed by the length by hemisphere by site interactions, was consistent with the length-related differences being centro-parietally distributed over the left scalp, but frontally distributed over the right scalp sites. This topographic distribution is not consistent with the distribution of the parietal positivity.

The length effects in this late region, which were present in the first presentations and not abolished by repetition, follow the same pattern as that for the 360-500 msec region. The 4-way interaction of length, repetition, hemisphere and site is reproduced in this window, and has a similar character. The interaction was apparently related to the diminution of the length differences being more marked at the frontal sites over the left hemisphere, but larger at the parietal sites over the right hemisphere. This interaction is the only indication that these late word length effects might be less robust to the effects of multiple repetition than the length effects found in the 230-360 msec window, where there was no interaction of any kind between length and repetition for the repeated traces.

These commonalities in the pattern of length effects over the 360-500 and 500-800 msec latency windows imply that the 2 windows captured the same length effects. This interpretation accords with the waveform differences depicted in Figures 6.2 to 6.5. There appears to have been a sustained positivity present in the 7 letter word traces relative to the 4 letter word traces throughout the later 2 windows.

The fact that this sustained positivity did not interact with repetition in such a way as to absent itself from the repeated waveforms was obviously unexpected. Unexpected both from the analyses undertaken in Part One and from the perspective, explored in the introduction to this chapter, that length effects would be expected early in the processing sequence. The presence of these late length effects in the repeated traces (leaving aside their high-order interaction with repetition) constrains the possible interpretations which can be made of the consequences of the experiments in Part One.

Length effects *are* present in that part of the waveform which the quantitative modelling technique suggested as the major locus of their action (the 230-360 msec region), but they are also present in regions of the waveform where their action would not have been suspected (the late region). Other linguistic factors were the major factors in the later regions of the waveform.

There are a number of possibilities concerning the apparent inconsistency between the results of this experiment and the results of the experiments in Part

One. First, it might be that length and frequency interact in the late part of the epoch in such a way that when there is variance related to each factor present, the length effects are abolished. This could be established, naturally, by a factorially designed experiment (with a large number of conditions to accomodate factors of length, frequency and repetition). Second, it is possible that word frequency is a more powerful factor than word length (i.e. gives rise to greater amplitude differences and more reliable effects) in the late part of the epoch. Information germane to this possibiltiy could be derived by undertaking a factorially designed experiment of similar design to the present one.

If this second possibility were found to be true, consistency would be preserved between Part One and Part Two by the conclusion that in some circumstances the quantitative modelling technique only identified the factor which gives rise to the greatest variance. This interpretation, and its consequences for the conclusions which can be drawn from the quantitative modelling approach taken in Part One, are further discussed in Chapter 8.

Relation of the results of this experiment to the quantitative modelling experiments in Part One

These results allow the qualified conclusion that most aspects of the results of this experiment which could have been predicted from the results of the quantitative modelling experiments have been borne out. There were indeed ERP differences present for repeated words which differed in length. These length-related differences were not abolished by multiple repetition.

The 230-360 msec word length effects predicted from the time course of the length effects in Part One were present. There was no suggestion that these effects were modulated by multiple repetition. By contrast, the later word length effects, which took a different form to the early effects, were involved in a high-order interaction with multiple repetition. The fact, however, that these late length effects were not completely abolished by multiple repetition can be consistent with the findings of the experiments in Part One if word frequency is found to be a more potent factor in the late region of the epoch. This gives a clear expectation about the outcome of the word frequency experiment, to the testing of which I now turn.

Chapter Seven

Word frequency and multiple repetition as determinants of the modulation of ERPs in a classification task

Introduction

Relation of the present experiment to the quantitative modelling studies in Part One:

This experiment was designed to test the conclusions, drawn in Part One, concerning the way in which word frequency modulates the ERP. The studies in Part One involved evoking a waveform to each of a small set of words. Considerable quantitative information had been generated about the words' psychological or linguistic properties (Word frequency, letter frequency, phonology, orthography, semantic similarity etc.). The extent to which variance between the words' waveforms was explained by the words' linguistic dimensions was then assessed.

The waveforms were evoked by presenting each stimulus repeatedly (50 times) in a random presentation sequence. One finding was that when subjects were called upon to make a decision concerning the category membership of presented words, despite the multiple repetition of the stimuli, the modelling technique uncovered variance sensitive to word frequency late in the epoch.

This frequency effect took the form of a parietally maximum positivity, peaking at about 600 msecs, which was large in response to individual words of low frequency and small to words of high frequency. The presence of a word frequency effect after the subjects had seen the stimuli perhaps a hundred times in a two hour session (which included collection of several sets of similarity judgements) implied that word frequency and familiarity could not be simply related, and was, from that perspective, rather surprising: low frequency words did not acquire the characteristics associated with high frequency words by being repeated within an experimental context.

Rugg (in press), however, has reported an apparently identical phenomenon; a parietally maximum positivity peaking around 600 msec, which was present for (once) repeated low, and absent for (once) repeated high frequency words.

In order to fully validate the findings of the quantitative modelling studies, though, 2 further features must be present. First, it requires to be demonstrated that after **multiple** repetition, frequency effects should not only take the form of a P600 present to low frequency items, but also be confined to this late portion of the waveform because variance sensitive to frequency was not found elsewhere. Hence, the early frequency sensitive effects apparent in the repeated waveforms of Rugg (in press) should be diminished by further repetition, while the late positivity should not. Second, the late word frequency effects, centred on the P600 deflection, should be greater than the late word length effects (as far as comparison~~s~~ between the experiments is possible, see Chapter 6). This is in order to account for the facts that frequency was the major factor late in the epoch and that word length effects were not observed in the latency range of the P600. Failure to establish these expected features in the present study would reduce the credibility of the results of the quanititative modelling technique, while finding all of them would constitute corroboration.

Behavioural studies of word frequency and repetition

The effects on performance in classification tasks of both word repetition and word frequency have been extensively documented. With respect to word frequency, the ubiquitous finding is that both the rapidity and accuracy of speeded responding is enhanced the greater is the frequency of occurrence of a word in written language (Forster and Davies 1984; McCann, Besner and Davelaar 1988; Monsell, Doyle and Haggard 1989; Norris 1984; Rubenstein, Garfield and Millikan 1970; Scarborough, Cortese and Scarborough 1977; Whaley 1978). With respect to word repetition, the most typical finding is that repeating a word within the context of an experiment enhances the speed and accuracy of classification relative to unrepeated words (e.g. Scarborough et al 1977; Monsell 1985).

The robustness of these effects has made them prominent features of many theoretical accounts of word recognition. Despite their centrality, however, the locus of their action is still in dispute. For example, one class of model imputes the frequency sensitivity of the information processing system to perceptual systems. More particularly this imputation is to processes by which the identity of a stimulus word is generated from its visual representation; that is, to processes of lexical access (e.g. Becker 1976; Forster 1976; McClelland and Elman 1986; Morton 1969). Another group of models suggests that the effects of frequency are mainly derived from task-specific decisional or response processes (e.g. Broadbent 1967; Morton 1968, 1982; Treisman 1971).

One aspect of the debate between these two types of model has been the extent to which frequency effects are obtained in every task which involves lexical access. To the extent that they are, models which suppose that access is itself frequency sensitive gain support. A series of recent studies, however, with behavioural data (Balota and Chumbley 1984, 1985; McCann and Besner 1987; McCann, Besner and Davelaar 1988), have militated against a purely access locus, by presenting evidence that frequency effects are not always present in data derived from tasks which (are alleged to) necessitate lexical access. Nonetheless, the empirical considerations on which these challenges are based are open to contention (Monsell et al 1989).

Similarly, there is dispute about the locus of repetition effects. This is principally because there is evidence that the way in which the repetition prime is processed is an important determinant of its efficacy in enhancing the speed of response to the target. When subjects were unaware of the repetition of the prime, because it appeared in a context different to that in which speeded responding was required (Oliphant 1983), recent exposure to a word did not necessarily influence processing of repeated items.

For some, these and other findings (Feustel, Shiffrin and Salasoo 1983; Forster and Davies 1984; Jacoby 1983), militate in favour of the possibility of a non-lexical mechanism, an episodic memory process, underlying the repetition effect.

In general, repetition and word frequency modulate reaction time interactively (Forster and Davies 1984; Norris 1984; Scarborough et al 1977; cf., Humphreys, Besner and Quinlan 1988). The benefits of repetition are typically greater for low than high frequency words, so that after repetition the effects of frequency are substantially less than for unrepeated items (though frequency effects are still detectable (Forster and Davies 1984)), a finding which has been called the "frequency attenuation" effect (ibid). The finding of interaction could be taken to imply (Sternberg 1969) that repetition and frequency share a processing stage. Issues concerning the locus of these two effects are consequently not independent (Humphreys et al 1988).

Electrophysiological studies of word frequency and repetition

A number of studies using event-related potentials (ERP's) have explored the modulation of brain potentials by these variables. Rugg (1985) recorded ERP's in a word-nonword classification (lexical decision) task. When items were repeated, the waveform evoked by the second presentation was found to be more positive than that for the first presentation in a latency region from 300 to 600 msec post-stimulus. This basic feature of the ERP repetition effect, a greater relative positivity in the waveforms of repeated items, appears in the subsequent studies even when no overt response is required of the subject to the conditions of interest (Rugg 1987).

The repetition effect has been found to depend on the lag between prime and target (Nagy and Rugg 1989). Immediate repetition, either of words or nonwords, sometimes evoked an early negative deflection which was absent to repetitions which occurred after several intervening items (Nagy and Rugg 1989; Rugg 1987), although the 300-600 msec effect was more robust to changes in inter-item lag (Nagy and Rugg 1989). This mid-latency effect, however, depended on the task required of the subject (Bentin and McCarthy 1989; Rugg and Nagy 1987; Rugg, Furda and Lorist 1988).

These results have been interpreted as evidence that early ERP repetition effects may be related to rapidly decaying prelexical processes, possibly those involving the identification of letters (which is conceived to be common to words and nonwords) (Nagy and Rugg 1989). The later repetition effect, by contrast, is interpreted as being determined primarily by postlexical task or decision-related processes (Bentin and McCarthy 1989; Rugg et al 1988), and as probably reflecting the diminution of a relative negativity in the waveform evoked by first presentation (Bentin and McCarthy 1989; Rugg et al 1988).

A smaller number of studies have explored the effects of word frequency on event-related potentials. Polich and Donchin (1988) investigated the effects on P300 of a word frequency manipulation across a range of probabilities. P300 latency was affected by frequency independently of probability, so that low frequency words evoked the component later in time. This was interpreted as evidence that "the frequency with which a word occurs in the language accelerates the initial stages of word processing, before the response production stages" (Polich and Donchin 1988 p 40).

Smith and Halgren (1987), in a rather complicated design, observed that, on initial presentation, the frequency effect took the form of a relatively greater negativity, maximal at about 400 msecs, to low frequency items (A similar feature appears in the waveforms of Polich and Donchin 1988). On repetition there was an enhanced positivity in the traces of both rare and common words, but no frequency effects were detected. Rugg (in press), however, noted that both the above studies employed paradigms in which response-related components were present in the traces, and suggested that these features may tend to obscure variance related to the experimental factors. Using a paradigm which confined the necessity to respond to occasional nonwords, a frequency sensitive negativity similar to that seen in both the above studies was observed to initial presentations. On repetition, however, there were early, frequency related, repetition effects (circa 250 msec) and a late positivity present only for low frequency words.

A fourth study exploring the effect of word frequency, by contrast to the above, found no effects of this variable in a delayed typicality judgement task (Stuss, Picton and Cerri 1988). Several effects related to frequency, however, approached significance. It was suggested that the elusiveness of frequency could be due to both typicality and frequency having their effects in the same direction (Stuss, Picton and Cerri 1988) (see the Discussion section of this chapter).

The results of these studies are not consistent. Three experiments (e.g. Polich and Donchin 1988; Rugg, in press; Smith and Halgren 1987) show effects of frequency while one (Stuss et al 1988) does not. Of the two studies that explored the interaction of frequency with repetition, one (Rugg, in press) found frequency related variance after repetition and one (Smith and Halgren 1987) did not. The present experiment therefore attempts to clarify the effects of word frequency and repetition on the evoked potential and to elucidate the possible functional loci of these effects.

The paradigm used by Rugg is more likely to enable variance related to frequency after repetition to be detectable than one in which response is made to the stimuli of interest (Rugg, in press). Hence the present experiment employs a variant of this task. This also ensures that the task structure is as close as possible to that in the category decision experiments in Part One. Also in the interests of preserving compatibility with the quantitative modelling experiments, it employs a category membership decision task and several repetitions of the words. This latter manipulation should serve to dissociate those features of the ERP repetition effect which may be due to "surprise" on initial repetition, from those which are related to other processes which are engaged by repeated stimuli, and to examine the predicted effect of multiple repetition on the diminution of effects at different latencies.

Issues of experimental control in a conventionally designed and analysed 'across-stimulus' experiment

Control for extraneous factors in a conventionally designed 'across-stimulus' experiment can only be accomplished by careful item selection. In Part One, I found that the major factors to which variance was sensitive in repeated single words were word length, word frequency and semantic similarity. Rugg (in press) controlled for initial letter, word length and included only open class words. Polich and Donchin (1988) selected on the basis of word length and included only nouns and verbs.

Accordingly, it was decided to effect control in this study for word length and for a general semantic factor, operationalised along the open-closed class distinction, and to assess possible item selection problems by including a second set of words as selection controls (see Bentin, McCarthy and Wood 1985).

Method

Subjects: Twelve young adults were payed for participation. Seven subjects were female, all reported themselves to be right handed, and all were naive as to the purpose of the experiment. None was being prescribed medication at the time of their participation. Five subjects had taken part in an evoked potential study previously.

Stimuli: As discussed above, control was effected by item selection for word length and for a general semantic factor by including only open class words. Forty items in both levels of frequency were selected on the basis of their being recorded only once in the Kucera and Francis (1967) corpus (low frequency), or having counts above 100 (high frequency). The two sets of words were balanced for word length so that each set contained a mean word length of 4.83 letters and a range in both sets between 3 and 7 letters. The two sets of words are appended in Appendix 5. These words constituted the main part of the design and were presented 4 times (see below).

A further set of 40 low and 40 high frequency items were chosen from the Kucera and Francis (1967) corpus to serve as item selection controls. Low and high frequency control words were selected according to the same criteria as for the items in the main part of the design. These words did not repeat.

Repeated items were inserted at lags taken from a rectangular probability distribution of median 10, and range between 6 and 14 intervening words, to minimise the predictability of where in the presentation sequence repetition might occur. The target words were unrepeated (to avoid the task becoming easier as the experiment progressed) animal names, inserted at random into the lists with a constant probability density ($p = 0.17$). The animal names varied from very low frequency (TAPIR) to high frequency (DOG).

Two lists were constructed with the serial positions of low and high frequency words reversed, to control for order of presentation or sequence effects.

Design and task: Subjects were asked to make a button press response with their preferred hand as quickly and accurately as possible when they detected an animal name. The nontarget words consisted of two levels of frequency (LF and HF) crossed with four levels of repetition (1st presentation and 3 repetitions, R0, R1, R2 and R3).

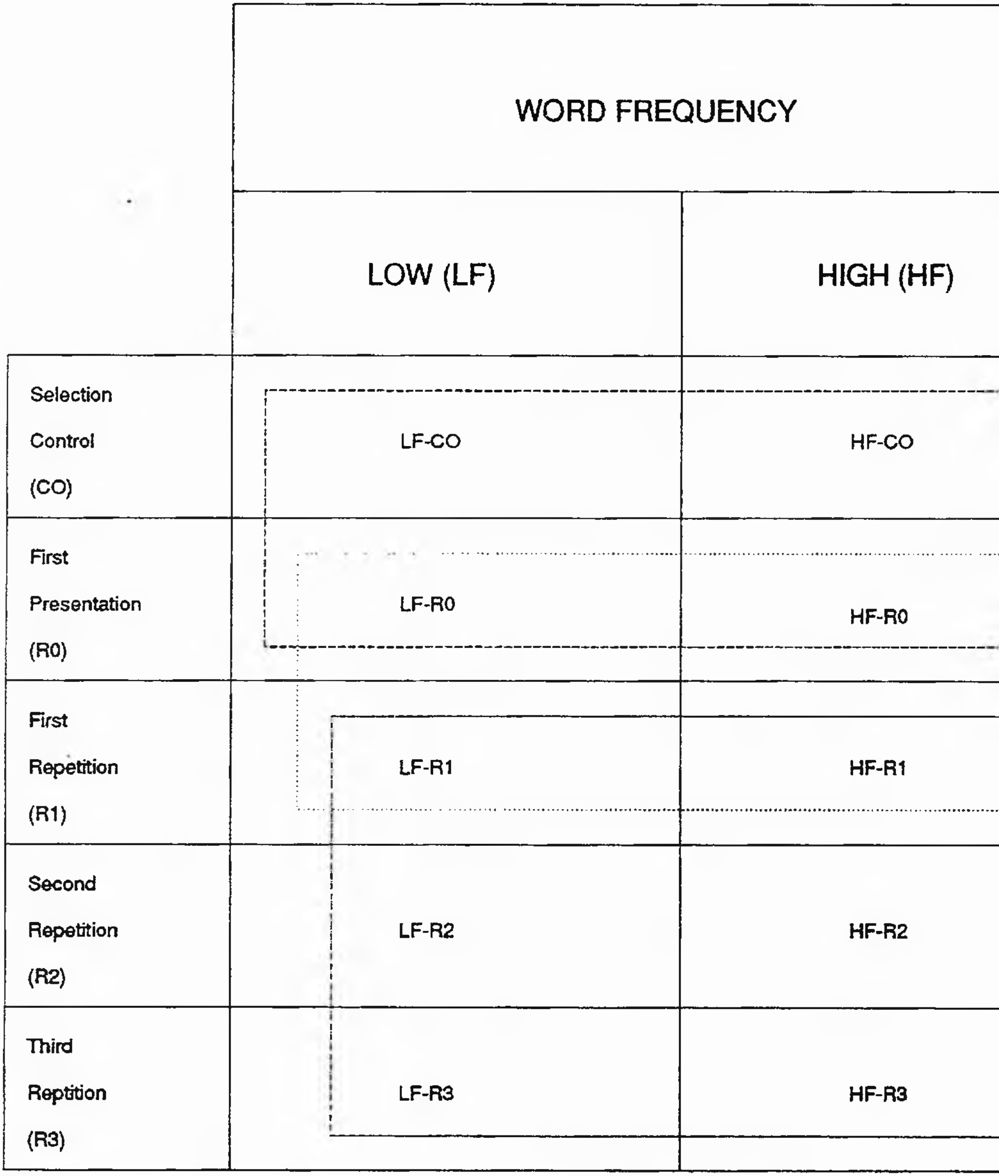
Control for item selection was achieved by a subsidiary part of the design which crossed two levels of frequency with two levels of word type, the one being the 1st presentations of words which would repeat and the other being the control items.

The design is set out in Figure 7.1. The two lists in which the order of presentation of the low and high frequency words were reversed were counterbalanced across subjects.

ERP recording:

All aspects of the electrode preparation and placement, amplification and digitization, stimulus presentation and averaging were identical to those detailed in Chapter 4.

Features of the design and analysis of experiment 7



	WORD FREQUENCY	
	LOW (LF)	HIGH (HF)
Selection Control (CO)	LF-CO	HF-CO
First Presentation (R0)	LF-R0	HF-R0
First Repetition (R1)	LF-R1	HF-R1
Second Repetition (R2)	LF-R2	HF-R2
Third Reptition (R3)	LF-R3	HF-R3

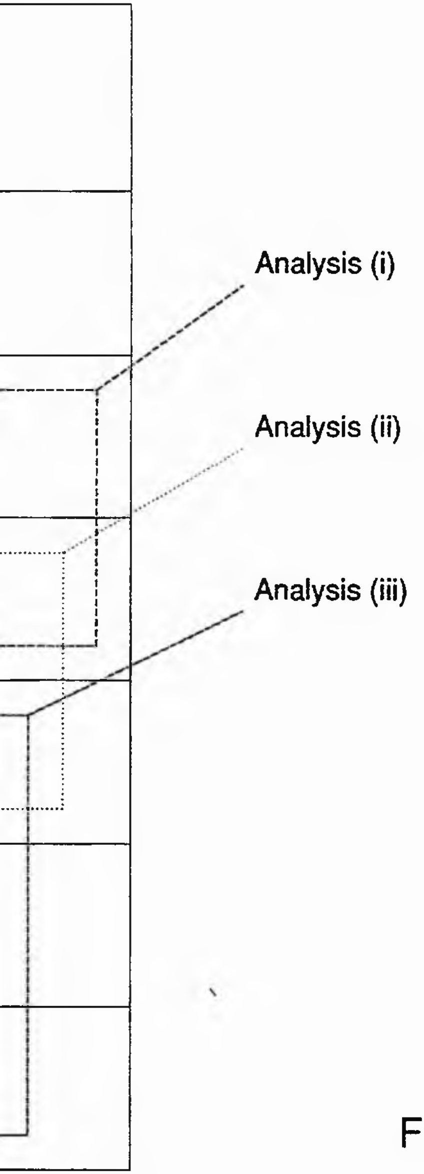


FIGURE 7.1

Procedure: After application and impedance testing of the electrodes, subjects were seated in a soft upright chair facing a table on which were fixed two response buttons at right and left, and a TV monitor. The task was explained to be that of responding as quickly and accurately as possible to animal names, which category was to include anything in the animal kingdom.

It was explained that words would be presented on the screen one at a time but that the targets would be mixed with non-target words as distractors. It was explained that some of the non-target words would repeat, but that this was irrelevant to the task of responding to the animal names. The reason for the repetitions was explained to be laziness on the experimenter's part: repeating non-targets made it possible to choose fewer words when designing the experiment. Subjects were also asked to try to withold blinking during each trial.

A sequence of 25 practice trials was administered, to ensure that the subjects were comfortable, understood the task and had no problem seeing the words, after which subjects had the option of altering their position or the lighting level. The practice trials were made up of presentations of a list derived from the target words from experiments 4 and 5, in which some repeats and animal names were inserted. None of the words in the practice sequence was repeated in the experiment.

When the subject was ready the experiment proper began, which was composed of 4 blocks of 120 trials, between each of which there was a short break. Subjects were debriefed during removal of the electrodes.

Results

Behavioural data: Mean Reaction time to the animal name targets was 701 msec, with a standard deviation across subjects of 84 msec. The mean error rate for misses (errors of omission) was 4.5%, and the mean false positive rate was 0.4%. These error rates were considered too low to submit to further analysis.

ERP Waveforms

Grand average ERP's are shown in figures 7.2 to 7.5.

Figure 7.2 shows the waveforms evoked by the first presentations of both high and low frequency words. The trace relating to the 1st presentation of the low frequency words (LF-R0) is deflected relatively more toward the negative in a latency window beginning at approximately 260 msec, than is the trace evoked by high frequency first presentations (HF-R0). This frequency difference peaks just after 400 msec and has a centroparietal distribution, with a larger amplitude and greater extension in time over the right hemisphere electrode sites.

Figure 7.3 shows the waveforms evoked by the first presentation of both high and low frequency words, together with the waveforms evoked by their first repetitions. Repeated words are characterised by greater positivity than is apparent in the first presentation waveforms, as has been demonstrated previously (Rugg 1985; Rugg et al 1988).

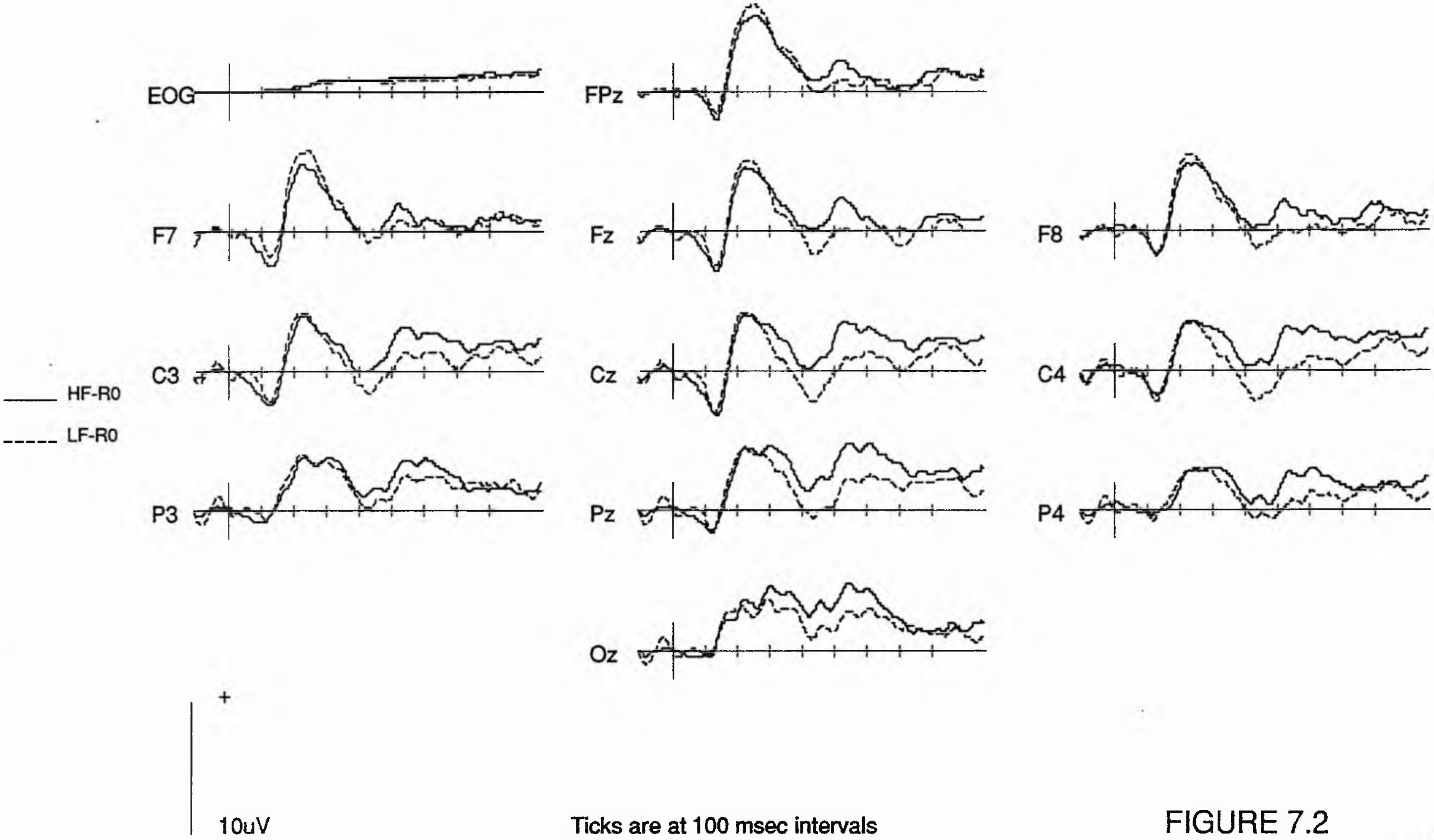

EOG
FPz
F7
Fz
F8
C3
Cz
C4
P3
Pz
P4
Oz
HF-R0
LF-R0
+
10uV
Ticks are at 100 msec intervals

FIGURE 7.2

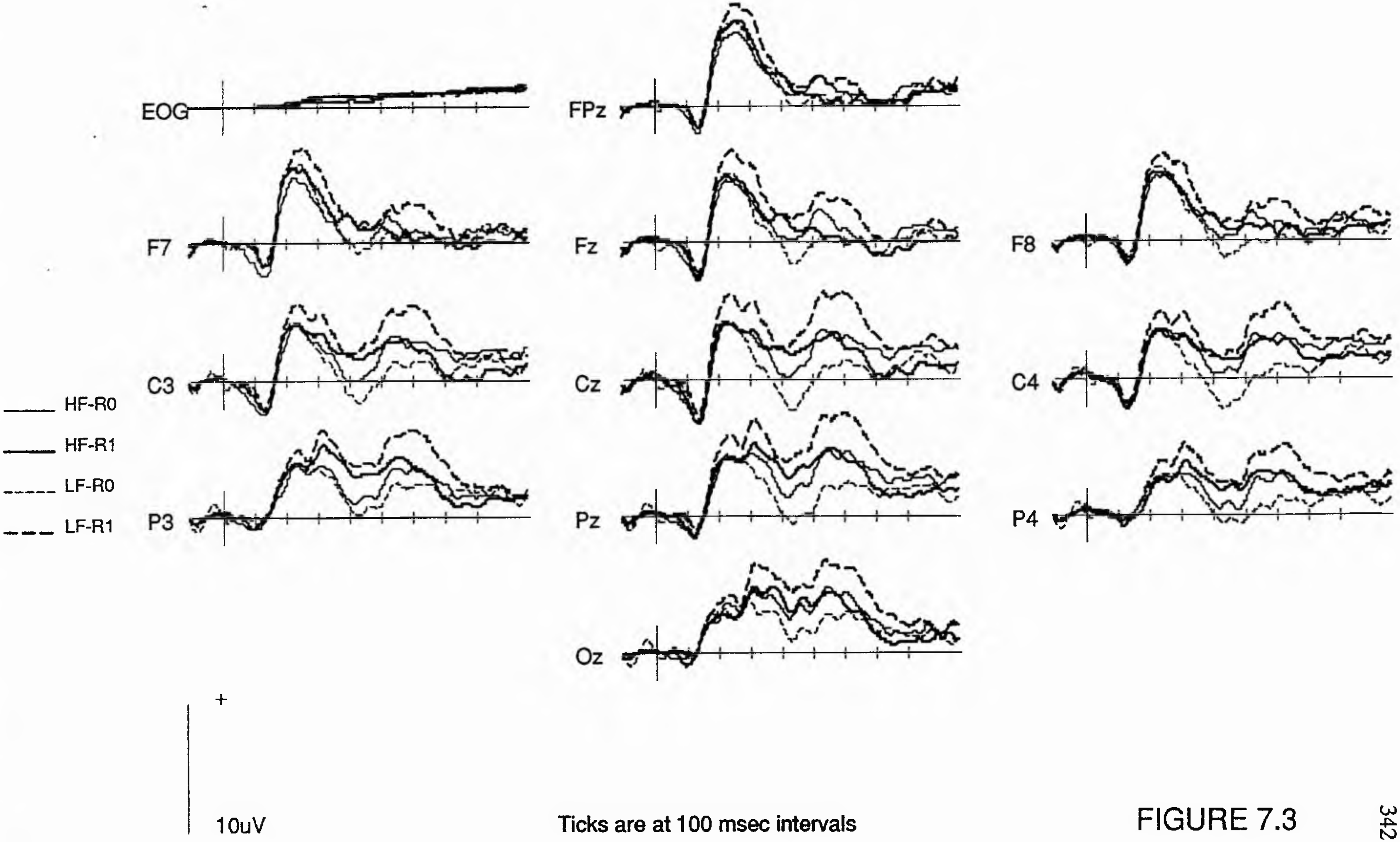
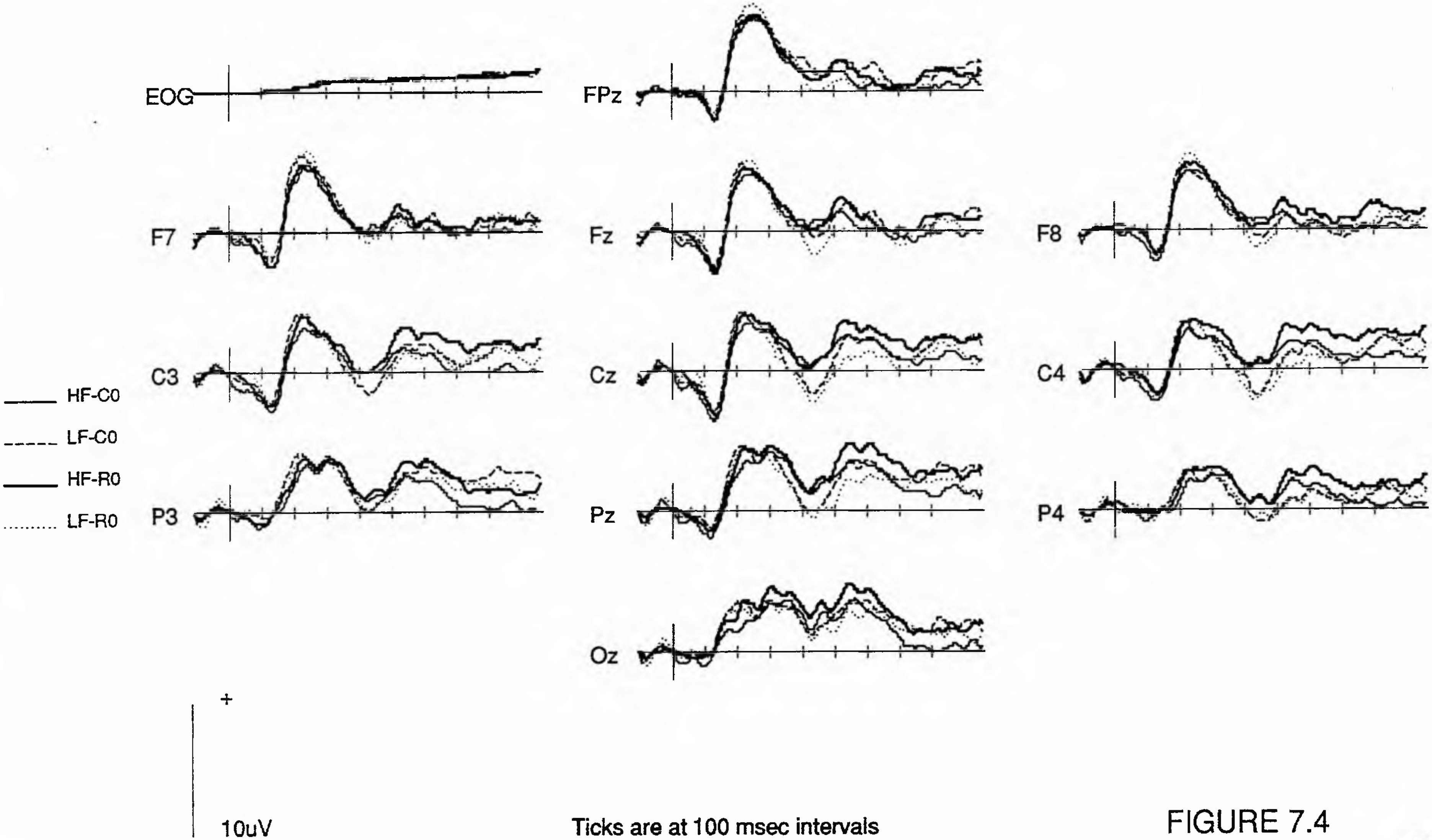
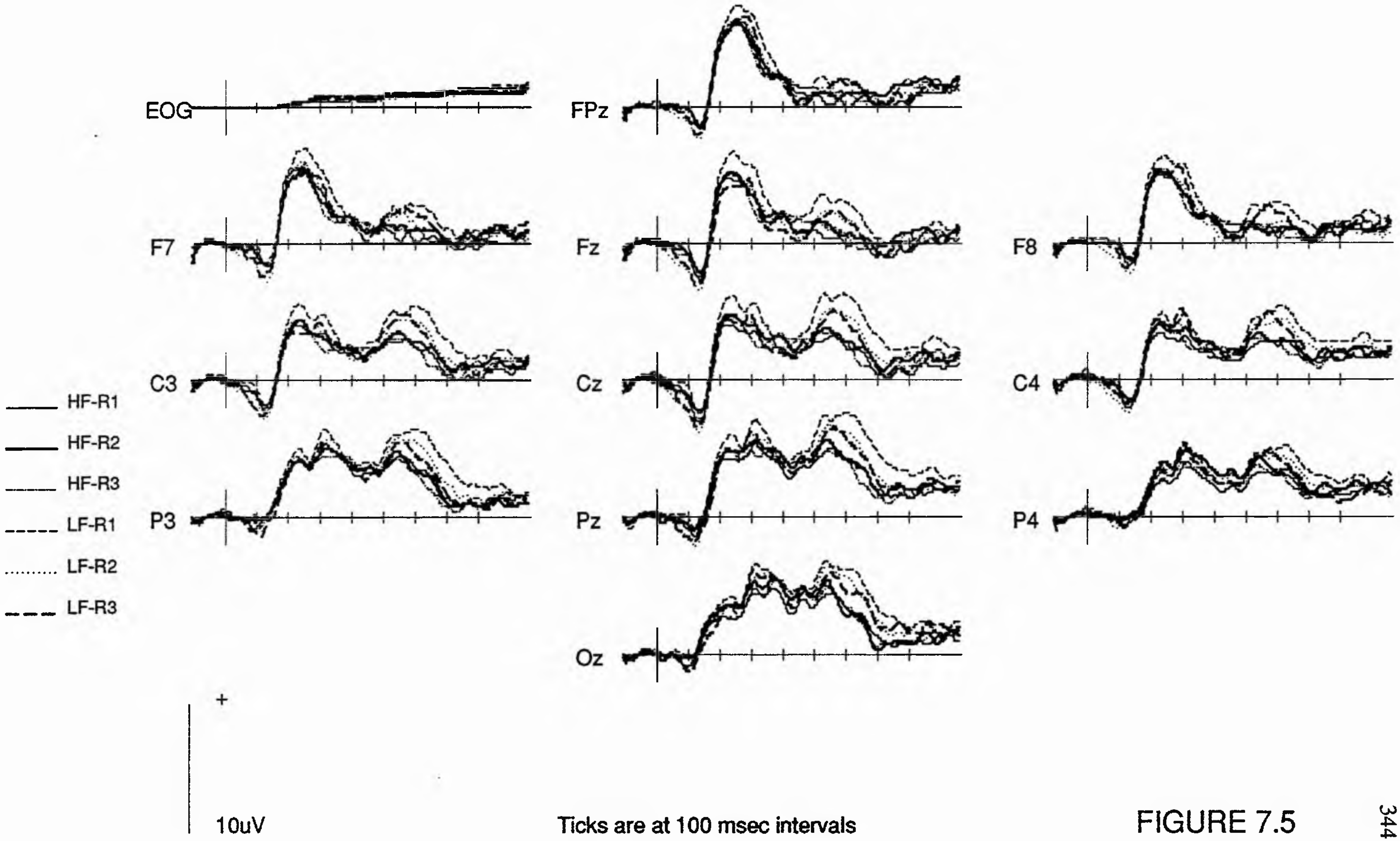

EOG
FPz
F7
Fz
F8
C3
Cz
C4
P3
Pz
P4
Oz
HF-R0
HF-R1
LF-R0
LF-R1
+
10uV
Ticks are at 100 msec intervals

FIGURE 7.3



Ticks are at 100 msec intervals

FIGURE 7.4



Ticks are at 100 msec intervals

FIGURE 7.5

The high frequency repetition condition (HF-R1) begins to diverge from the HF-R0 trace just prior to 300 msec and this relative positivity is sustained until approximately 500 msec, apparently consistent with previous interpretations of ERP repetition effects as the diminution of a 'default' negativity in the first presentation waveforms (Rugg et al 1988).

The low frequency repetition waveform (LF-R1), however, begins to diverge from the LF-R0 trace rather earlier, at about 180 msec, and has a more complex morphology. A large positivity, possibly reflecting the influence of two modulations, is present between 180 and 350 msec, at which point the amplitude of the ERP is similar to that evoked by the HF-R1 condition. A large parietally maximum positivity peaking at about 600 msec characterises the LF-R1 waveform in the latter part of the epoch.

Figure 7.4 shows waveforms evoked by the HF-R0 and LF-R0 conditions and the high and low frequency item selection control conditions (HF-CO and LF-CO, respectively). There is little difference between the traces on the basis of word type. The control waveforms can be seen to reflect the same centroparietally, right hemisphere distributed frequency-related negativity mentioned in relation to the first presentation waveforms, HF-R0 and LF-R0.

Figure 7.5 shows the waveforms evoked by the three repetitions of low and high frequency words (LF-R1, 2 and 3; HF-R1, 2 and 3). While the low frequency first repetition trace (LF-R1) may be slightly more positive going than waveforms derived from subsequent repetitions, the 'P600' evoked by repeated low frequency words is still plainly apparent as a difference between the LF traces and those from HF repetitions.

ERP analyses

The data were measured by taking the mean amplitude of the waveforms relative to the 100 msec pre-stimulus baseline in selected latency regions. Time windows of 200-300, 300-500 and 500-800 msec were chosen both to capture the apparent differences in the waveforms and to preserve compatibility with other studies (e.g. Rugg, in press). Measurements in these time bands were submitted to repeated measures analysis of variance (ANOVA), with Greenhouse-Geisser correction of degrees of freedom (Keselman and Rogan 1980).

Analyses in each latency region were undertaken separately for midline and lateral sites, in order to explicitly examine the data for the presence of hemispheric asymmetry. In the midline analyses, 3-way ANOVAS were performed for 3 sets of factors.

Analysis (i) included factors of word type (control vs first presentations), word frequency and electrode site, in order to assess the adequacy of item selection control and to examine the effect of the frequency manipulation on both control and R0 conditions.

Analysis (ii) employed factors of repetition (R0, R1), frequency (LF, HF) and electrode site. This analysis was undertaken in order to examine the effects of the frequency and repetition manipulations on initial repetition, and to be directly comparable with aspects of Rugg (in press).

Lastly, analysis (iii) used factors of repetition (R1, R2 and R3), frequency (LF, HF) and electrode site. This analysis examined the effects of the frequency manipulation and multiple repetition on the repeated conditions.

In the lateral site analyses, 4-way ANOVAS were conducted with the same sets of factors as for the midline analyses, but with the additional factor of hemisphere.

A technical problem during one recording session caused the loss of signals from the P4 electrode in one subject. A missing data procedure was performed which involved substituting the average of all other subjects' data at this site. Table 7.1 shows the mean amplitudes of each condition over selected electrode sites.

200 - 300 msec

There were no main effects or interactions in the analysis which targetted the first presentations of the control items and the items which would repeat (analysis (i), see Figure 7.1).

Table 7.1

	High Frequency					Low Frequency				
	Fz	Cz	Pz	C3	C4	Fz	Cz	Pz	C3	C4
200-300										
Control:	4.0	3.4	3.9	3.1	3.7	4.4	3.7	4.5	3.7	3.0
R0:	4.4	4.0	4.7	3.7	3.7	4.7	3.9	4.4	3.8	3.3
R1:	4.8	4.1	4.8	4.0	3.5	6.6	6.1	6.3	5.4	4.9
R2:	5.0	4.9	5.4	4.2	4.2	5.3	4.9	5.6	4.4	4.4
R3:	4.9	4.5	5.1	4.0	3.9	5.2	4.9	5.6	4.3	4.5
300-500										
Control:	1.3	1.6	2.8	0.9	1.2	0.8	0.0	1.7	-0.1	-0.3
R0:	1.4	1.7	3.1	1.3	1.6	-0.1	-0.6	1.1	-0.2	-0.8
R1:	1.8	2.7	4.4	2.5	2.2	3.0	4.0	5.7	3.4	2.9
R2:	2.1	3.3	5.0	2.8	2.9	1.9	3.1	5.1	2.6	2.5
R3:	2.1	3.3	4.8	2.7	2.7	1.7	3.0	5.1	2.6	2.6
500-800										
Control:	-0.1	1.7	2.5	0.8	1.6	0.8	2.0	3.1	1.6	1.4
R0:	1.0	2.9	4.0	2.7	3.0	-0.1	1.1	2.3	1.3	1.1
R1:	-0.3	1.8	3.4	1.6	2.3	1.8	4.6	6.0	3.9	4.3
R2:	0.8	2.5	3.6	2.3	2.7	0.9	3.3	4.5	2.9	3.5
R3:	0.3	2.0	3.0	1.7	2.3	0.9	3.3	4.8	2.9	3.8

The analysis (ii) ANOVA, however, which focussed on the first presentation and first repetition conditions of both high and low frequency words, gave a main effect of frequency (F1, 11 = 5.22, p = 0.04).

There were no main effects or interactions in the analysis which focussed on the 3 repeated conditions (analysis (iii)).

300 - 500 msec

Analysis (i) ANOVAs on the data derived from the 300-500 msec epoch, which examined word frequency effects and the adequacy of item selection control on unrepeated conditions, did not reveal any word type effects. These analyses, however, produced a main effect of frequency (F1, 11 = 4.98, $p < 0.05$). In addition, there was an interaction between frequency and site (F2.2, 24.7 = 9.38, p = 0.001) due to the centro-parietal sites being more negative for the low frequency words. For the lateral sites, there was a frequency by hemisphere (F1, 11 = 8.82, p = 0.01) interaction, indicative of greater negativity for the low frequency words over the right hemiscalp, and a hemisphere by site interaction (F1.6, 18 = 6.47, p = 0.01), due to the hemisphere differences being more marked over the centro-parietal sites.

The main effects and interactions of the ANOVA which examined the first presentations and one repetition (analysis (ii)) were as follows. There was a highly reliable effect of repetition (F1, 11 = 26.50, $p < 0.001$). There was an interaction between frequency and repetition (F1, 11 = 21.37, p = 0.001) because of a greater increase in positivity on repetition for the low frequency words than for the high frequency items. There was an interaction of repetition by site (F1.1, 12 = 6.07, p = 0.03) which indicated that the effects of repetition were larger at the centro-parietal sites. There was an interaction of hemisphere with site (F1.7, 19.1 = 9.47, p = 0.002) indicative of the fact that hemispheric asymmetry was largest at the centro-parietal sites.

[An analysis on all four levels of repetition gave essentially the same pattern of results as the analysis (ii) reported above, but with a pattern of interaction sufficiently interesting to be worth reporting. This analysis revealed a frequency by hemisphere (F1, 11 = 7.97, p = 0.017) and a hemisphere by site interaction (F1.6, 18 = 6.9, p = 0.008). The repetition by hemisphere and the frequency by site variance ratios did not reach significance (both F's < 1.3). These results have some implications which are discussed in the discusion section of this chapter.]

There were no main effects or interactions related to the experimental manipulations in the analysis (iii) ANOVA, which examined the effects of frequency and multiple repetition on the repeated conditions.

500 - 800 msec

The 500-800 msec epoch analysis (i), which examined first presentations of both control and to-be-repeated items, revealed a frequency by hemisphere interaction (F1, 11 = 10.66, p = 0.008) indicative of the extended influence of the right hemisphere dominant frequency sensitive process which peaked around 400 msec. There was no main effect of word type.

Analysis (ii), based on 2 levels of frequency and the first and second presentations, showed the following effects. There was a main effect of repetition (F1, 11 = 7.11, p = 0.02) and a main effect of site (F1.7, 18.8 = 6.27, p = 0.01), due to the presence of a large positivity with a parietal maximum. There was also a highly reliable frequency by repetition interaction (F1, 11 = 26.81, p < 0.001), due to a much greater increase in positivity at this latency for the low frequency words. A repetition by site interaction was present (F2.2, 24.2 = 4.58, p = 0.02), due to the effects of repetition being larger at the centro-parietal electrodes. There were interactions of hemisphere with repetition (cf., the 300-500 msec analyses) (F1, 11 = 5.49, p = 0.04) due to asymmetries taking the form of greater negativity over the right hemisphere for the first presentations, but greater positivity being present over the right hemisphere after repetition. There was an interaction between hemisphere and site (F1.7, 18.5 = 3.78, p < 0.05) due to hemispheric asymmetry being largest at the parietal leads. Finally, there was a 3-way interaction between frequency, repetition and site (F1.6, 17.9 = 5.16, p = 0.02) indicative of the fact that repetition brought about a much greater increase in positivity for the low frequency words at the parietal electrodes.

The analysis based on repetitions only (R1, R2, R3; analysis (iii)) revealed no effect of repetition, nor any interaction of repetition with frequency. This analysis did, however, reveal a highly reliable frequency main effect (F1, 11 = 17.16, p = 0.002) and an effect of site (F1.6, 17.8 = 6.97, p = 0.01). In addition, there was an interaction of frequency with site (2.1, 22.6 = 8.55, p = 0.002), reflecting the presence, in the low frequency but not in the high frequency traces, of a parietally maximum positivity. Finally, there was a 3-way interaction between repetition, hemisphere and site (F2.4, 26 = 3.39, p = 0.04) which reflected an equipotential effect of repetition over the right hemisphere, as compared to a parietally maximal distribution of repetition effects over the left hemisphere.

[Following Rugg (in press), an analysis based on high frequency conditions only (HF-R0, HF-R1, HF-R2 and HF-R3), showed that in this period the first presentation and subsequent repetition-evoked traces did not differ (F2.7, 29.9 = 1.364, p = 0.27).]

There were no effects of word type in any analysis of any time window.

Discussion

One major feature of the results of this experiment is that word repetition and word frequency modulate ERP's in an interactive fashion. The simplest interpretation of the presence of repetition by frequency interaction in a behavioural variable would be to suppose that repetition and frequency both act on the same locus, in functional terms, in the processing system (Sternberg

1969). This interpretation, however, if unqualified, scarcely does justice to the complexity of these data.

For example, in the 300 to 500 msec time window, there was a highly reliable frequency by repetition interaction. But the frequency manipulation interacted with hemisphere and the repetition manipulation did not. Repetition interacted with electrode site while frequency did not. Hence an interpretation of this pattern of interactions is to suggest, contrary to the simplest account of repetition by frequency interaction (above), that these two factors modulate at least partially independent generators (see also Besson, Kutas and van Petten 1989; Rugg, in press).

Under the assumption that these ERP differences are related to the activities of processors which have a functional significance in the production of behaviour, the data appear to support the conclusion that there are at least three loci which are affected by these manipulations. These are the early effects, the 300-500 msec effects and the late positivity. These phenomena may be discriminated on latency and scalp distribution grounds and on the grounds that they may be differentially sensitive to the two experimental factors.

Early effects

The earliest analyses of variance, for the 200-300 msec period, showed that there was an early effect of frequency on the ERP. As reported in Rugg (in press), this effect was mainly mediated by the first repeat of the low frequency words. This effect was, however, not robust to further repetition since the repeated traces

ANOVA showed no effect of frequency. Hence the results of the early latency window analyses of this experiment are entirely consistent both with Rugg (in press) and the quantitative modelling experiments in Part One.

This early frequency-related phenomenon coincides in latency so closely with the P200 peak that it appears, superficially, to be an enhancement of the amplitude of P200 (see Figure 7.3). It is intriguing to ask, then, whether this effect recorded here is related to N1-P200 effects evoked in some semantic priming tasks, which have been interpreted as preconscious meaning attainment (Boddy 1981, Boddy and Weinberg 1981; Boddy 1986), or to the P200 effects seen in sentence processing experiments (Kutas et al 1988).

To address the first issue, a number of additional analyses with the same factor sets as before, were carried out on the latency region of N1 and on N1 to P200 peak to peak measurements. These indicated no effects related to the experimental manipulations. The N1-P200 analyses, however, consistently gave a main effect of hemisphere, reflecting a greater peak to peak measurement over the left scalp sites. As the studies reported by Boddy (Boddy 1981; Boddy and Weinberg 1981; Boddy 1986) showed no lateralisation for this measurement, and as the source of early language-related variance is apparently no earlier than P200 in the present data, the relation between the N1-P200 studies and the P200 effects evoked here is probably not close. The present data, however, are somewhat supportive of the idea that quite high level information relating to the presented words is processed surprisingly early (Boddy and Weinberg 1981). A processor which is sensitive to the frequency of occurrence of an item in English, a property which is not a physical aspect of the stimulus, appears to begin to modulate scalp recorded activity at about 200 msec.

In a variety of sentence processing experiments (reviewed by Kutas et al 1988) an asymmetry of P200 has been observed, in which the right hemisphere sites are more positive, particularly over occipital sites. The early frequency-sensitive effect in this study showed no tendency to be larger over the right, indeed the deflection was (non-significantly) larger over the left sites (see Table 7.1). Hence, no persuasive case can be made for relating the sentence processing P200 to the P200 effects evoked here. It may be that these manipulations have targetted only a subset, or a different set, of processors having their influence on the ERP at this latency, than are activated by sentence processing or semantic priming tasks. If so, it is possible that a process giving rise to a right hemisphere predominance at P200 is evoked by either or both of passive exposure conditions and/or sentence context: both the equipotential-P200 priming experiments (Boddy 1981; Boddy and Weinberg 1981; Boddy 1986), and the present study required a decision about each word of the subject, and presented rather isolated words when compared to the sentence presentation paradigm of Kutas and colleagues.

The early positivity declined from its amplitude on initial repetition with further repetition. Since, so far as is known, no previous study has systematically examined the effects of multiple repetition, it is not possible to identify this aspect of these phenomena with previously identified components. The early onset of this modulation, however, and the pattern of their relation to the experimental variables suggest~~s~~ an obvious functional interpretation.

In Part One I found that variance between ERP's evoked by individual words was best explained in a window from stimulus onset to about 300 msec by word length variables (see Chapter 5). After this point it was found that variance was sensitive to a semantic factor captured by judged semantic similarities. This change in the sensitivity of variance from a low level, possibly physical, property of the stimuli to a relatively high level semantic attribute was unimaginatively interpreted as reflecting the occurrence of lexical access.

The timing and nature of this result converge with a number of results from eye fixation experiments. These reveal that 300 msec is just short of the average time that a competent reader invests in a content word in fluent reading (Carroll and Slowiaczec 1986; Inhoff 1984; Just and Carpenter 1980; Rayner and Pollatchek 1981). The length of a fixation is argued to reflect the time taken to achieve recognition and to integrate word meaning into sentence context (Just and Carpenter 1980 (Although this must be treated with caution: Kliegl, Olson and Davidson 1982)). Application of regression techniques to fixation data has suggested that the process of identifying word meanings from their visual representations, for words of the length employed in this study, takes place in about the same time, 300 msec (Just and Carpenter 1980). On this basis, and as a rule of thumb, effects prior to 300 msec could perhaps be interpreted as reflecting processes prior to, or during, the activation of semantic attributes of the stimulus. Effects after this point could be imputed to post-lexical processes.

In this context, it is possible that this early effect reflected the activity of processors whose function is related to lexical identification. These processors are modulated particularly by low frequency words, are not apparent on initial presentation, are maximal on first repetition, and decline on further repetition such that multiply repeated low frequency words do not differ from high frequency words. These properties are rather similar to the properties proposed for a labile lexical mechanism sensitive to long term aspects of exposure (frequency) and to recent exposure (repetition) (Just and Carpenter 1980; Morton 1969; Becker 1976; Forster 1976).

Mid-latency effects

Both the control items, which did not repeat, and the first presentations of the words from the main part of the design showed a difference between high and low frequency words which took the form of a centroparietally distributed negativity which was larger for low frequency words. This onset around 260 msec and continued, over some sites, to 800 msec. This phenomenon is very similar to that reported in previous studies which have employed word frequency as a factor (Smith and Halgren 1987; Rugg, in press; cf., Stuss et al 1988). A similar feature is apparently present in the waveforms of Polich and Donchin (1988, Figure 2), but it was not the main focus of their interest in that study and was not reported. It may be, however, that their finding of a smaller P300 to low frequency words compared to high frequency ones, independent of the level of probability (of nonwords in a word-nonword classification task), is consistent with the presence of a long lasting relative negativity in the low frequency waveforms. Indeed, their principal component analysis indicates a larger 'N200' (peaking at 400 msec) for low frequency items which somewhat overlaps with the principal component identified as P300 (Polich and Donchin 1988, Figure 3).

None of the previous studies using frequency has identified a right hemisphere emphasis for this first presentation frequency difference. This was either because they did not record from lateral sites (e.g. Polich and Donchin 1988) or because the asymmetry was not present (e.g. Rugg, in press, Table 1; Smith and Halgren 1987).

The present study, by contrast, evoked waveforms with a clear right hemisphere dominance in the frequency effects for both control and R0 items. The reason for the absence of asymmetry in the foregoing studies and for its presence here is not known, but there is an obvious difference of task. The Smith and Halgren (1987) and Rugg (in press) experiments involved a word-nonword classification (lexical decision) task where this experiment employed a category membership classification task. It is possible that a task which lends more emphasis to the identification and manipulation of meaning associated with each stimulus, rather than its presence as an entry in some notional internal lexicon or recognition as a high frequency pattern (relative to nonwords), is more likely to engage lateralised processors. The sentence processing experiments reviewed by Kutas et al (1988) also reveal an asymmetry, at the same latency as the present frequency effect, across most of the extent of the sentences. These tasks may also lend emphasis to the identification and manipulation of the meaning of each word in the sentence, as the development, during presentation, of a meaning-enriched sentential context seems to be a natural consequence of reading attended text. Possible relations of the present asymmetry to asymmetries in data derived from studies with markedly different task structure, such as delayed phonological matching tasks (e.g. Rugg 1984) are yet more uncertain.

Quite apart from the reasons for the presence of asymmetry, this frequency effect, taking the form of a centroparietal, right hemisphere negativity between 300 and 800 msec, is not readily distinguishable on latency and scalp distribution grounds from the N400 evoked in sentence processing and semantic priming experiments (Kutas and Hillyard 1982, 1983; Kutas et al 1988; Bentin et al 1985). Is this feature, then, N400?

Traditionally, there have been two sets of criteria employed to determine the identity of components evoked in different studies (Gaillard 1988; Kutas et al 1988). One has stressed the interpretation of distributional and latency aspects as determinants of whether deflections are likely to share common neural generators. The other has weighted the experimental manipulations, such as task structure, and the putative cognitive processes assumed to subserve the functions required by the task. These two sets of criteria are here (as elsewhere, for example in the 'Is N400 an N200?' debate (e.g. Ritter, in press)) somewhat in conflict. Frequency is not straightforwardly a semantic variable, yet the ERP feature it evokes is not distinguishable from the 'semantic' N400 on morphological grounds, and hence might be presumed to be the signature of a very similar generator (or generators). Unqualified application of either set of criteria, however, is problematic. It may be that interactions between signals from concurrently active generators, and the transfer functions between active brain tissue and scalp, muddle the real distinctions between generators, so that ERP's are insufficiently precise to distinguish different processors in different tasks (cf., Rugg, Kok, Barrett and Fischler 1986). Equally it may be that models borrowed from microdomains of information processing psychology are ill equipped to capture the underlying commonality of processors engaged by different tasks or experimental manipulations. Perhaps a weighted combination of both sets of criteria is required, as well as some caution (Kutas et al 1988).

On morphological grounds there seems little ground for the suggestion that the frequency effect evoked here and the 'semantic' N400 are the result of distinct generators. As far as the more manipulation-related criteria are concerned, there are at least two possibilities.

First, perhaps it is the case that the N400 is not evoked only by semantic factors, but by some more general factor, presently unidentified, which can encompass semantic and word frequency aspects, as well as other aspects, some of which are not even ostensibly linguistic (see Osterhout and Holcomb 1989; Rugg 1984; Stuss, Sarazin, Leech and Picton 1983). A second possibility is that the N400 is only, or most particularly, sensitive to semantic factors, but the word frequency manipulation carries with it one or more semantic variables.

With respect to the latter possibility, word frequency has been shown, for some corpora, to covary with semantic concreteness, number of meanings (polysemy) and imageability (Gernsbacher 1984; Rubin 1980; Whaley 1978). Imageability, which is not a semantic variable per se, is closely associated with the semantic variables of 'ease of predication' (Jones 1985) and of meaningfulness (Paivio, Yuille and Madigan 1968). Hence there may be several semantic factors carried by word frequency. An account of the frequency effect at mid latencies in terms of N400 could therefore be given which does little violence to the interpretation of N400 as primarily an indicant of semantic processes. Such an account, however, leaves open the first possibility noted above, that the negativity is sensitive to frequency per se, bearing in mind that the covariance of frequency with these other semantic factors is not high (across the Whaley (1978) and Rubin (1980) studies and across the variables of imageability, concreteness and meaningfulness, the association is on the order of (Pearson) $r = 0.14$), and that other features of the waveforms related to the frequency manipulation, such as the P600, are rather dissimilar to those evoked by semantic and priming related manipulations. On both morphological and cognitive grounds, though, my suspicion is that this feature is indeed N400.

The interaction of frequency and hemisphere, and of repetition with site in the 300-500 msec region implies that the scalp distributions of the frequency and repetition induced modulations are dissociable. This is not to say, however, that the present data necessitate reappraisal of the notion that repetition effects are constituted by a diminution of a 'default' N400 present in first presentation waveforms (Bentin and McCarthy 1989; Nagy and Rugg 1989; Rugg et al 1988).

This process of attenuation may indeed be one source of variance in the waveforms evoked here, but it is not sufficient to account for all the effects which the analyses have detected. To the extent that the repeated traces stayed at approximately the same amplitude at this latency, regardless of multiple repetition (see Figure 4), it is plausible to believe that the most salient difference in this latency window is the presence of an N400 in the first presentation traces and its absence in the repeat traces. This pattern of interaction of the frequency and repetition variables with scalp distribution is very similar to that in a study examining the relation between semantic priming and repetition (Besson et al 1989), in which a somewhat asymmetrically distributed priming modulation interacted with a more equipotential repetition effect. This concurs with the conclusion, above, in relation to the distribution of first presentation frequency effects: in terms of their attenuation by repetition, the frequency effects seen here may be quite closely related to previously documented 'semantic' effects in the latency region of N400.

The data militate against the interpretation of the repetition-induced reduction of 'N400' as due to an overlay of P3 (see also Besson et al 1989; Rugg et al 1988). While there was a late positive component with a parietal maximum present in the repeated low frequency traces, the repeated high frequency waveforms did not evoke any greater positivity in the same time period, indeed they did not differ from the first presentation trace. In this respect the N400 attenuation effect and the presence of a late positivity are here dissociated: there was no late positivity for repeated high frequency items but these words nonetheless showed an identifiable repetition effect.

It has recently been suggested that N400 is not modulated by semantic priming when the prime is masked (Brown, Hagoort and Swaab 1989). It would be interesting to know whether the present N400-like frequency effect is also abolished by masking the stimuli. This would entail a similar paradigm to that of Brown et al (1989), taking care, however, not to follow their conflation of the critical comparison with a difference of task (e.g. semantic judgement vs. lexical decision), and could illumine the relation of this frequency effect to lexical and post-lexical processes. This is unfortunately outside the aims of the present focussed set of studies.

Late effects

The repeated low frequency conditions evoked a parietally maximum positivity which peaked around 600 msec. This was a frequency sensitive process in that repeated high frequency items showed no such late positivity. The P600 in the low frequency traces was not abolished by multiple repetition and so appears to reflect the action of a late, and rather stable, frequency sensitive processor.

It may be that this feature is only manifest in the waveform when a decision is required of the subject: the first quantitative modelling experiment, using a passive exposure paradigm (experiment 1, Chapter 3), produced neither a late positivity nor any evidence for frequency effects anywhere in the epoch. In order to further discriminate the conditions for evoking this feature, it would be interesting to employ a conventional paradigm to establish whether a decision (reflected in a silent counting task) or memorisation is sufficient to evoke this effect.

It is possible that the P600 is related to performance in recognition memory tasks as it is apparently insensitive to the degree of intra-experimental exposure which a word receives. Similarly Kinsbourne and George (1974, experiment 1), exploring the advantage of low frequency words in recognition memory tasks, found that their frequency effect was not due to a change in effective frequency after experimental exposure (cf. Smith and Halgren 1987), evidencing some stable frequency related process. This concurs with the subjective intuition that repeating a low frequency word in a limited context (e.g. prosaic, prosaic, prosaic...) does not abolish the knowledge that it is, in fact, a low frequency word.

Although the possible relations of this positivity to memory processes, and the reason for its absence when subjects were passively exposed, are interesting issues to pursue, they are somewhat outside the focus of the present studies. Discussion of the possible mnemonic correlates of this late word frequency effect is contained in Rugg (in press), to which the reader is referred.

On the functional loci of frequency and repetition effects

These results are supportive of the view that there are several loci sensitive to the effects of word frequency and repetition. These include, on the present interpretation, both lexical and post-lexical mechanisms. Two aspects are, however, cold comfort to theories which suppose that access processes are the only, or the major, locus of frequency effects.

The first is that the most stable frequency sensitive effect (P600) is too late to be imputed to recognition processes, and may only be evoked when a decision is required. This implies that there exists a post-access processor which is engaged by particular task demands and characterised by a non-labile frequency related organization. The second is that unrepeated words both show the greatest effect of frequency (Forster and Davies 1984), and evoked the N400 feature. If the N400 is related to effects on performance, and is itself post-lexical, as has been argued (Brown et al 1989; Rugg, in press), then the most salient locus of frequency effects may also be post-lexical.

Relation of the results of this experiment to the quantitative modelling experiments in Part One

These results allow the conclusion that every aspect of the results of this experiment which could have been predicted from the results of the quantitative modelling experiments has been borne out. There is indeed a late positivity present for repeated low frequency items, peaking at about 600 msec, which is not abolished by multiple repetition. The early and mid-latency frequency effects are, by contrast, abolished by multiple repetition such that the repetitions of low and high frequency words differ only in respect of the presence of the P600. Multiply repeated words do not evoke frequency sensitive variance elsewhere in the epoch.

This experiment yielded qualified support for the expectation that word frequency would be a more effective factor in the late part of the epoch than was word length (see Chapter 6). The comparison of the repetition only analyses in the 500-800 msec latency window of the 2 experiments is instructive. Despite the smaller number of subjects in the frequency experiment, and hence its lower 'power', the variance ratio for the frequency manipulation was more than double that for the length experiment. The amplitude difference associated with each manipulation, at the electrode site at which each was maximal, was 1.8 uV at Pz for the frequency effect but only 1.3 uV at Oz for the length effect (corresponding to a 30% difference).

These findings support the conclusion, suggested tentatively in Chapter 6, that the quantitative modelling experiments did not show length effects late in the epoch because of the presence of a more powerful linguistic factor. Hence there need be no inconsistency between the findings of the studies in Part One and those in Part Two. The logical consequences of the conclusion that the configurational analyses detected only the *major* source of variance, when there was an imbalance between the influence of two or more factors in any latency window, for the conclusions which can be drawn from these studies is discussed in Chapter 8.

To summarize the present experiment, the surprising finding of a late word frequency effect in ERPs evoked under conditions of multiple repetition has been corroborated by this conventionally designed experiment. The expectation of this result was generated by my application of a quantitative modelling rationale to the structure of differences between single word ERPs. This consistency between the exploratory across-stimulus studies in Part One, and the confirmatory studies in Part Two, suggests that the former converged on linguistic factors which do, indeed, underlie differences between ERPs in these paradigms.

Chapter Eight

General Discussion

An overview

I begin this chapter by reviewing the issues which I hoped to address (see Chapter 1), the way in which I chose to address them (see Chapter 2) and the results of the empirical investigations (see Chapters 3 to 7). This review is intended to aid assessment of the degree to which these studies have succeeded in informing the intended issues.

The central question with which these studies have been concerned was that of which linguistic factors, if any, give rise to ERP variance in these paradigms. The importance of this question was suggested to be twofold.

The first reason for seeking to investigate this area was related to methodological issues, particularly those concerned with experimental control. Only after those intrinsic dimensions of words which have a salience in producing ERP differences have been identified can a properly informed choice be made of linguistic material for list presentation experiments. The intention was to provide information for future studies, in which list items are not to be counterbalanced across experimental conditions, and to point up the degree to which experiments undertaken in the past, where items were not counterbalanced, were likely to have conflated different sources of variance.

The second reason for pursuing this line of enquiry was the importance of the question itself. Many studies in the area of ERPs and language processing have aimed at identifying the properties and relations embodied in verbal material which are related to ERP variance (e.g. Bentin, McCarthy and Wood 1985; Chapman, Bragdon, Chapman and McCrary 1977; Kutas and Hillyard 1984; Polich and Donchin 1988; Rugg 1987; Smith and Halgren 1987). The discovery of covariations between the linguistic and electrophysiological domains both may inform existing language processing models (but see below) and be a stimulus to the development of future construals of the brain processes which mediate language.

The methodology I employed to pursue these issues was based on a quantitative modelling approach. This approach was intended to bear a 'family resemblance' to the multiple regression studies conducted by psycholinguistic researchers, faced with the identical problem in the behavioural domain. Regression, however, was not readily applicable to ERP data. I had to develop a new quantitative modelling approach which was intended to be better tailored to the awkward properties of ERP methodology and data structure. The innovative character of the approach necessitated the testing of the results of Part One with conventionally designed and analysed experiments in Part Two.

The results of experiments 1 to 5, which constituted the quantitative modelling studies in Part One, suggested that there were three linguistic factors related to ERP variance in these paradigms. These linguistic factors were word length, semantic similarity and word frequency. These factors had their maximal effects at different periods in the epoch. Word length effects were maximal earliest, followed at 250-300 msec by meaning related differences. In the category membership decision experiments, a word frequency effect was present late in the epoch. These 3 linguistic factors were found to be poor predictors of ERP variance in 'inappropriate' regions of the epoch (e.g. the word length model explained very little variance late in the epoch).

Two empirical tests of the conclusions drawn in Chapter 5 concerning the action of these linguistic factors on the ERP were conducted. The word length experiment (Chapter 6) indicated that there were, indeed, word length effects present in ERPs derived from repeated words at just the latency range predicted. The experiment, however, also indicated that there were sustained, small but reliable differences between ERPs to repeated items of different length late in the epoch. This feature of the results was not expected on the basis of the results of Part One, where only word frequency effects were found to be present at that latency.

The word frequency experiment (Chapter 7) indicated, among other things, that there were, indeed, word frequency effects in ERPs derived from repeated words late in the epoch. Word frequency effects at other latencies were found to be abolished by multiple repetition. The word frequency effects were mediated by the P600 deflection. All features of the results of this experiment were exactly as predicted from the results of the quantitative modelling experiments. In addition, it was found that the word frequency effects in the late part of the epoch were both more reliable and larger than the word length effects.

The one inconsistency between the results of the quantitative modelling studies and the results of the factorially designed experiments was the presence of late word length effects in the latter. There may be a number of ways construe this unexpected feature of Part Two. The interpretation which, I believe, accounts best for all the facts is that the quantitative modelling technique only detected the **major contributor** to ERP variance in some circumstances.

These circumstances probably existed when two or more factors were active, but when one was much more effective in giving rise to ERP variance than the other(s). This feature of the data analysis strategy could be seen in Part One. For example, both word length and semantic similarity effects were present at particular latencies in the category membership decision experiments, but the whole epoch analyses showed only word frequency effects. Hence it is possible that if word length effects were present in the late region of the epoch of these experiments, they were rendered undetectable by the presence of a word frequency effect large enough to override them in the late analyses; large enough, indeed, to override the effects of all other factors in the whole epoch analyses.

What inferences can be made from the results of these exploratory studies?

There is an asymmetry, I believe, between the positive and negative aspects of the results of these studies. The positive aspects of the results, the findings of the 3 linguistic factors which modulated the ERP waveform and the latencies at which each was the major factor, seem quite well supported. But the negative aspects, such as the findings that the other linguistic factors which were tested against ERP variance did not modulate the waveforms, must be interpreted with even greater caution.

The first issue here is that the finding of 'no effect' can be due either to the absence of the effect or to the insensitivity of the analytic technique. This problem exists for any analysis strategy. This is particularly important in the interpretation of these studies, though, for the following reason. I drew the conclusion above that, in some circumstances, the quantitative modelling technique probably only identified the factor which gives rise to the greatest variance in a particular latency window. In the early, mid-latency and late windows, word length, semantic similarity and word frequency were the major influences on the waveforms. It cannot be concluded that variance related to the competing linguistic factors was not present in the ERPs, only that such variance was substantially less than that due to the 3 identified factors.

The second issue is the obvious proviso that all the ERPs in Part One were evoked by repeated words. The possibility cannot be excluded that some of the factors which did not explain variance well, would have done so if the ERPs were evoked by first presentations only. If, however, the same pattern of influence of linguistic factors on the ERP is sought in future (to aid experimental control, perhaps), the stimuli could be repeated a small number of times. The near absence of effects of multiple repetition, over and above the effects of initial repetition (see Chapters 6 and 7), suggests that a similar pattern of effects could be derived more economically, in terms of the numbers of repetitions.

The consequence of this asymmetry between the positive findings and the findings of no effect is that it would not be wise to conclude that the factors which were not identified as ERP variance predictors can be safely excluded from consideration when selecting items for future experiments. It can only be concluded that their contribution to ERP variance was much smaller, in these experiments, than the contribution from the 3 identified factors.

Another consequence of the fact that the data analysis approach in Part One may only have detected the 'largest' factors is that it was probably naive to suppose a simple succession in the relation of ERP variance to each of the 3 identified factors (see Chapter 3). The evidence was that the effects of the factors may have overlapped in time. For example, the suggestion from the furniture term experiments (experiments 4 and 5) was that word length effects were present in the mid-latency analyses, as were the effects of semantic similarity (see also Chapter 5). Similarly, it is likely that effects of word length were still present in the waveforms during the 400 to 800 msec window, in which word frequency effects were dominant.

The dominant linguistic factor, nonetheless, in the early part of the epoch was word length and the dominant factor in the mid-latency part of the epoch was semantic similarity. While there was evidence that these factors influenced ERP differences in a temporally overlapping fashion, there was a change in the *dominance* of the 2 factors. To the extent that the sensitivity of ERP variance changed from physical features to semantic aspects, it may be that these differences were related to neural systems subserving the function of establishing a word's meaning from its visual representation. This change occurred at what might be regarded as the right latency for the process of lexical access; 250-300 msecs (see Chapter 7). This conjecture could be examined by assessing the degree to which manipulations thought to delay lexical access (e.g. degradation of the stimulus) also delay the variance sensitivity change observable in this type of experiment.

It was not possible to use a conventionally designed experiment to examine the effects of semantic similarity, because suitable lists could not be constructed. The evidence for the presence of this factor from the quantitative modelling experiments is only as persuasive as the extent to which the modelling technique was validated by conventional designs using known analytic techniques. The experiments in Part Two found that those aspects of the data which could have been predicted from Part One were present as expected. To this extent the technique used to generate these expectations seems to have been accurate, and there seems little reason to doubt, therefore, the veracity of its results in relation to a factor which was not explorable by conventional means.

Some general issues

Keeping in mind the fact that these results are the products of **exploratory** studies, there are several issues which they seem to provoke.

The paradigm employed in experiment 1, which involved attended but passive exposure to verbal material was very similar to one of the tasks employed in a recent study using positron emission tomography (PET). Peterson, Fox, Posner, Mintun and Raichle (1988) measured emission from subjects' brains in a subtraction paradigm involving several levels of task. The simplest task involved simply fixating a cross-hair. The next task involved the attended but passive exposure to a series of single words. This was followed by a third task which required the subjects to speak the presented words aloud, and a fourth which required that a semantically appropriate verb be spoken to each presented noun. It was assumed that "each level of task ... include[d] all information-processing components present in the lower levels" (Peterson et al 1988 p 167). Blood flow was then measured over a 40 second period during which a radiolabelled bolus of water was distributed over the brain. The pattern of blood flow in each level of task was then subtracted immediately preceding task, to focus on the information transactions supposed to be specific to that task.

The interpretation of this type of subtraction experiment requires a number of assumptions. Not only must it be assumed that the addition of putative extra processing stages leaves processes targetted by simpler tasks intact, but the interpretation of each subtraction is based on assumptions of what processes each task entails. One assumption was that passive exposure to verbal material excites no more than word form codes (Peterson et al 1988).

This assumption, on the basis of the results of the experiments reported here, particularly experiment 1, seems unlikely to be true. In a paradigm which required the subjects to attend, but which required no active semantic processing, variance sensitive to semantic aspects of words was detected. Indeed, in the passive exposure experiment this sensitivity was sustained in time. It is possible that techniques which lack good resolution in the time domain, such as PET, would be more, and not less, likely to detect brain regions involved in semantic processing in passive exposure conditions than in tasks which involve the generation of a response. As a consequence it may be that imputation of the representation of 'word form codes' to the extrastriate regions which were activated by the passive exposure task of Peterson et al (1988) is qualifiable.

Variance sensitive to several precise aspects of presented words has been identified in brain signals derived from the synchronous activations of millions of neurones (Wood 1987). It seems reasonable to briefly explore why this might be the case.

The finding of correspondences of this character between aspects of language and signals emitted by the organ which mediates them (see also Kutas and Hillyard 1984) confutes the expectation that "only the grossest correspondences can be found" (Fodor 1981 p 135). This expectation is related to the notion that the representation of a word, at every level, is a recursively defined syntactic structure (Fodor and Pylyshyn 1988). That there should be relations between neural mass actions and word *meaning* is particularly peculiar from this perspective, as the basic assumption regarding the representation of meaning is that there exist expressions whose form (i.e. their syntax) mirrors their semantics. This finding is peculiar because there are no relations obtaining between the formal statements or data structures in a formal program running on a computer, and what the hardware 'looks' like when it runs (How could there be? The same program may have quite different implementations). In the same way, it is a remarkable coincidence that electrical signals indicative of 'what the wetware looks like' when language processing takes place, should so precisely, and unnecessarily, mirror the supposed formal properties of the expressions whose manipulation is assumed to mediate language.

I mentioned in Chapter 3 that a relation between ERPs evoked by exposure to individual words and the structure of their judged semantic similarity was expected in advance of the facts. The conceptual route by which this expectation was developed, briefly, began with a dissatisfaction with the account of representation remarked on above. The notion of expressions whose form constrain their meaning faces a simply stated but very serious difficulty. A prototypical example of an expression is an argument. For every argument of valid form which proceeds from true premises to true conclusions, there is an infinity of arguments of identical valid form which proceed from false premises to false conclusions. The *form* of an expression cannot constrain its meaning because of the distinction between validity and truth/falsity (see Flew 1975).

This difficulty, together with the absence of a single example of an expression whose syntax reflects its meaning (the provision of which would refute the above), underline the difficulty of giving an account of the relations between representations and their objects in the framework of formal relations (see also Churchland 1986; Searle 1980). These relations between representations and their objects are sometimes called 'intentional relations' (Searle 1983), because of the property, called Intentionality, of 'aboutness' or 'ofness' which representations possess. For example, my representation of the computer on which these words are being written is 'about' or 'of' or 'directed toward' or 'fitted to' the computer. If intentional relations were unlikely to be formal relations, the only other possibility (that I could think of) was that they might be causal relations.

Now, causal relations do not seem promising candidates for an approach to intentional relations, particularly if one has a 'billiard ball' model of causation (intrinsic Intentionality from atoms banging in the void?). But consider the following 2 analogies. The population of antibodies deployed in an immune response just is 'about' or 'of' or 'fitted to' the configuration of antigens on invading bacteria. The population of organisms which inhabit a particular habitat just is 'about' or 'of' or 'fitted to' the ecology which has selected for them over time. Both these examples capture something like the sense of 'aboutness' required, are both obviously examples of causal relations and are both examples of selective systems (operating by clonal selection and natural selection, respectively).

Several researchers have explored the idea that the brain might be a selective system (Changeux and Danchin 1976; Edelman 1978; Jerne 1967; Young 1973). On a selective account, the representational states of the nervous system are distributed activity patterns over populations of neurones.

In this kind of account, the semantic similarity of two representations, as judged by a subject, will be a function of the commonality of two implicated populations, as will be the probability of recognition confusion between them (see also Ashby and Perrin 1988). By assuming that the transfer functions between brain tissue and scalp are not so nonlinear as to disturb a monotonic relation between active populations and the signals to which they give rise (i.e. the more similar are two spatially heterogeneous populations, the more similar will be their electrical signatures at the scalp), I made the following prediction. It was predicted that it would be possible to recover a structure of differences between ERPs evoked by exposure to individual words which should have the same statistical structure as that for their judged semantic similarity.

It is an historical fact that the analytic strategy developed here was originally designed expressly for the purpose of subjecting this prediction to empirical test. The results of this series of studies were supportive of the idea that it was, indeed, possible to recover variance related to semantic similarity. On balance, however, this finding provides only the weakest support for populational approaches to neural representation. First, the correspondence between ERPs and semantic similarity could be due to any of a number of semantic factors which happened to be well captured by the semantic similarity judgements which subjects made of the words (see Chapters 3 to 5). Second, the processes of electrogenesis of even the well characterised ERP components are poorly understood, and so no elaborate claims can be made concerning the electrogenesis of the mid-latency semantic effects which did not appear to correspond to known ERP components (see Chapter 5). In this respect ERPs were not an appropriate methodology for approaching these issues.

The assumption of a monotonic relation between activity patterns across cortical populations and electrical signals at the scalp, however, could be tested by the means I have developed here. Testing would involve searching for ERP correlates of differential firing in *empirically identified* cortical populations, rather than in *hypothetical* populations involved in semantic processing. For example, Georgopoulos and his coworkers (e.g. Georgopoulos, Kalaska, Caminiti and Massey 1982; Georgopoulos, Kalaska, Crutcher, Caminiti and Massey 1986) have identified a population of cells in the primary motor cortex of the Macaque which are differentially activated when movements in different directions are made. The activity of the whole sampled population (represented by a 'population vector') is closely related to the direction of arm movements. It would be possible to assess the degree to which these populations are detectable at the scalp by recording movement-related potentials in the same task. A global ERP variance model derived from all the sampled movement directions could then be fitted by models derived from end-point coding or direction coding hypotheses, to assess whether ERP variance is related to the underlying cellular population dynamics. This approach might usefully inform one of the several unknowns of ERP electrogenesis.

A method of general utility in cognitive neuroscience?

Attempts are constantly being made in cognitive neuroscience to uncover correspondences between levels of brain and behaviour (e.g. ERPs and RT, or ERPs and the properties of cell populations), particularly between data at the different levels. The quantitative modelling approach used here to examine possible relations between psychological variables and ERPs might be a useful exploratory tool in this area.

The utility of the method would be subject to a number of provisos. One proviso is that some pre-existing knowledge of at least one of the levels which are to be investigated is required. The statistical control effected in this approach is accomplished by generating models for the possible ways in which the stimuli (or responses) differ. Those factors which are responsible for driving variance at a particular level would be identifiable by the high variance explained statistics for models representative of these factors. Factors which did not explain variance well could not then be invoked to explain the differences in whatever dependent variable is chosen.

Another proviso is that it was very useful in these exploratory studies to employ 2 different sets of stimuli. The difference between the 2 groups of words in the interrelations of the linguistic factors which they embodied was particularly useful to the task of deciding which factors were more likely to have been responsible for ERP variance in the case of two models both being well fitted to the empirical variance. This proviso reflects the fact that the application of a quantitative modelling methodology does not absolve the experimenter of responsibility for stimulus choice.

An example of the application of this methodology was given above in relation to the proposal to test the assumption of monotonicity in the transfer functions between brain tissue and scalp. Another example, which is very close to the application explored in the present work, might be as follows. Heit, Smith and Halgren (1988) describes a population of cells in human medial temporal lobe which are selectively excited by particular words or faces. But what is it about these stimuli that excites them? For the words, does the cell that fires to "LUCK" like four-letter words, or words rhyming with "SUCK" or words that mean "FORTUNE" or some other aspect of this word, such as its occurrence in English. That is, is it particular word-lengths, phonologies, meanings or word-frequencies (etc.) to which the cell is tuned?

The question is answerable by presenting a set of words for which models representing each probable dimension have been erected. By treating each cell's responses as a variable (or data source), a quantitative model of variance between words could be generated, and the degree to which this model is fitted by the 'psychological' models computed.

Functional significance of effects and some remarks on a general approach to cognitive neuroscience

There are frequent attempts in the literature of cognitive psychophysiology to relate aspects of ERP data to information processing models of cognition. These attempts to relate the two domains appear to have 2 motivations. The first is the hope that ERP phenomena may inform theoretical issues in cognitive psychology. The second is the hope that the integration of the two domains may aid the interpretation of ERP effects, which, in the absence of a theoretical framework, may only be treated phenomenologically.

There are, however, some problems associated with the process of identifying ERP components or ERP differences, as for any neurobiological data, with particular putative processing stages of information processing models. These models are intended, almost universally, as formal and functional accounts of the behaviour in their domain, and are operationalised in terms of behavioural variables (see Broadbent 1958; Fodor 1981; Mehler, Morton and Jusczyck 1984; Pylyshyn 1984).

To try to relate any domain of empirical data to models which are formal and functional accounts of cognition is also implicitly to accept distinctions essential to those accounts. Characterisation of cognition in such terms, as anticipated above, specifies no aspect of the instantiation of the supposed processes in neural systems. In just the same way that a formal program specifies no aspect of what the hardware might 'look' like when the program runs, there are no entailed relations between a formal cognitive model and the electrical characteristics of the neural systems which may or may not instantiate it.

Formal models may only be related to neurobiological data on cognition by the invocation of auxiliary hypotheses. These auxiliary hypotheses either specify the nature of neural systems which subsume the characteristics of the formal model in the real brain, or identify aspects of neurobiological data with elements of the putative computational process.

These auxiliary hypotheses leave open the opportunity for the formal modeller, not wishing to have his psychological theory disproved "by some irrelevant physiological research" (Broadbent 1958), to challenge the linking hypotheses rather than his own position. It has been suggested that there exists an imbalance between the ability of purely behavioural data to decide between models (Kintsch 1980) and the 'power' of these models to resist refutation (Johnson-Laird, Herrmann and Chaffin 1984), such that every information processing psychologist may have his own theory (Watkins 1984). If this is true, it is hard to see how cognitive neuroscience will mitigate the problem. This perspective suggests a pessimistic prognosis for the epistemological problems of the formal approach to cognition.

In practice, the identification of an ERP component, or ERP difference, with a processing stage is made by assuming that deflections are related to systems which have some functional significance in the production of behaviour, or by trying to 'map the component in cognitive space', or by attempts to draw chronometric parallels.

The 'mapping in cognitive space' program does not seem to have been unequivocally successful (e.g. Verleger (and commentators) 1988). The chronometric program is uncertain in the era of parallel and distributed models of cognition, which render the possible timing of operations more variable than in the tractable era of additive factors. This is an especially poignant problem in the present case, as the plethora of information processing models purported to explain the action of the 3 factors which have been identified here are not even agreed on the serial order of processes: multiple activation models, for instance, allow the possibility that activation of high-level semantic attributes may occur prior to word identification (Monsell et al 1989).

For a variety of theoretical and practical reasons then, little attempt was made in the foregoing experiments to relate the pattern of results to existing information processing models. Instead speculations on the functional significance of the processors implicated by these studies involved a number of data driven considerations which gave rise to simple chronometric frameworks.

The application of even a rough chronometric system to ERP differences, though, has to deal with a further problem. Effects may be interpreted either as indicants of the activity of a processor which is concurrent with the difference in the traces (i.e. the waveform differences are 'signatures' of the activity of the processor to which they relate), or as indicants of the output of a prior but covert processor (i.e. waveform differences are the *effects* of a processor, perhaps on diffuse neuromodulatory systems). For example, on the former interpretation, the onset of N400 at 260 msecs is probably too late on chronometric grounds (see Chapter 7), for N400 to be imputed to processes of lexical access. On the latter interpretation, however, a covert processor prior to 260 msecs might well be implicated in such processes. In either case, of course, the processor related to the ERP effects cannot be later than the empirical phenomenon (Rugg 1987; Rugg et al 1986).

I think that there are 2 possible procedural strategies which might, in the long term, mitigate these problems associated with attempts to find theoretical frameworks in which to place ERP phenomena.

The first is to attempt to find parallels, not between neurobiological data and cognitive *models*, but between neurobiological data and *data* at the behavioural level. Some of the most striking successes in modern cognitive psychophysiology have taken the form of demonstrations of covariation between behavioural data and ERPs, even if this 'data to data' comparison was not explicitly the research strategy (e.g. Duncan-Johnson and Donchin 1977; Kutas and Hillyard 1984). This strategy underlay, in part, the choice of the methodology employed in the present studies. The hope, of course, in following this strategy is that this kind of approach will map out some of the empirical 'terrain' in advance of later theoretical construction. Demonstrations of covariation between the behavioural and neurobiological domains have, in the past, been helpfully suggestive of the types of processes to which an ERP phenomenon might relate (e.g. Kutas and Hillyard 1984).

The second strategy is much more ambitious. Rather than 'borrowing' models from cognitive psychology, the strategy is to develop models capable of specifying both the neurobiological processes which mediate behaviour, and the characteristics of the behaviour they mediate. This strategy will require a difficult and long term effort since neural systems are so exquisitely complex. It requires, furthermore, that we "have to do the work" (as the theorist Richard Feynman used to say). Nonetheless, the advances in theoretical and empirical neurobiology over the last decade prompt the hope that such a strategy may not be impossible. I am not convinced that there are many viable short cuts to tenable interpretations of the functional significance of ERP components.

Conclusion

I think that these exploratory across-stimulus studies may have made three substantive advances.

The quantitative modelling approach, which I have here applied to ERPs evoked by exposure to individual words, seems not to have badly misrepresented the variance present in the data. There are many opportunities to employ an approach of this nature in cognitive neuroscience. It is possible, therefore, that the method itself will be of utility.

The results of these experiments suggest the conclusion that there are 3 major factors driving ERP variance in these paradigms: word length, semantic similarity and word frequency. Experimental control should be effected for at least these 3 factors when selecting items for conventional experiments, where stimuli are not to be completely counterbalanced between conditions.

The difficulties associated with across-stimulus studies in event-related potentials were discussed in Chapter 1. These difficulties are most acute in the domain of single item ERPs. They are, indeed, so acute that the intrinsic linguistic factors which are embodied in individual words have not been investigated interpretably in any previous study. It may be that these exploratory across-stimulus studies are the first demonstration that systematic information can be derived about the linguistic character of words from single word ERPs.

Reference List

Abt K. (1981) Problems of repeated significance testing. **Controlled Clinical Trials 1:377-381**

Aine C. and Harter M.R. (1986) Visual event-related potentials to colored patterns and color names: Attention to features and dimension. **Electroencephalography and Clinical Neurophysiology 64:228-245**

Aldenderfer M.S. and Blashfield R.K. (1984) Cluster analysis. Sage University paper series on quantitative applications in the social sciences, Sage publications, Beverly Hills.

Alho K., Tottola K., Reinikainen K.,Sams M. and Naatanen R. (1987) Brain mechanisms of selective listening reflected by event-related potentials. **Electroencephalography and Clinical Neurophysiology 68:458-470**

Alvey (1980) GENSTAT: A general statistical program. Numerical Algorithms Group, Oxford.

Adrian E.D. and Matthews B.H.C. (1934) The Berger rhythm: Potential changes from the occipital lobes of Man. **Brain 57:355-385**

Arrezzo J.H.G., Vaughan H.G. and Koss B. (1977) Relationship of neuronal activity to gross movement related potentials in monkey pre- and postcentral cortex. **Brain Research 132:362-369**

Arthur D., Lewine J., Schmidt A., Oakley M., George J., Roeder S., Hillyard S., Aine C., Kutas M. and Flynn E. (1989) Event-related magnetic fields of the human brain during information processing. Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX), Noordwijk, The Netherlands

Ashby F.G. and Perrin N.A. (1988) Toward a unified theory of similarity and recognition. **Psychological Review 95:124-150**

Baddeley A., Conrad R. and Thompson W. (1960) Letter structure of the English language. **Nature 186:414-416**

Balota D.A. and Chumbley J.I. (1984) Are Lexical Decisions a Good Measure of Lexical Access? The Role of Word Frequency in the Neglected Decision Stage. **Journal of Experimental Psychology**: Hum. Perc. Perf. **10**, 340-357

Balota D.A. and Chumbley J.I. (1985) The Locus of Word Frequency Effects in the Pronounciation Task: Lexical Access and/or Production? **Journal of Memory and Language 24: 89-106**

Barrett G., Blumhardt E.L., Halliday A.M., Halliday E. and Kriss A. (1976) A paradox in the lateralization of the visual evoked response. **Nature 261:253-255**

Barrett S.E., Rugg M.D. and Perrett D.I. (1987) Event-related potentials and the matching of familiar and unfamiliar faces. **Neuropsychologia 26:105-117**

Bateman W.G. (1915) The naming of colours by children: the Binet test. **Pedagogical Seminary 22:469-486**

Battig W.F. and Montague W.E. (1969) Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. **Journal of Experimental Psychology Monograph 80:1-45**

Beaumont J.G. (1983) The EEG and task performance: A tutorial review. In: Gaillard A.W.K. and Ritter W. (Eds) Tutorials in ERP research: Endogenous components. North Holland.

Becker C.A. (1976) Allocation of Attention during Visual Word Recognition. **Journal of Experimental Psychology: Human Perception and Performance 2: 556-566**

Begleiter H. and Platz H. (1969) Cortical evoked potentials to semantic stimuli **Psychophysiology 6:91-100**

Belacicco (1977) On clustering time varying data. **Journal of the Royal Statistical Society**

Bentin S. and McCarthy G. (1989) Task Dependent Factors Influencing ERP Repetition Effects. Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX),, Noordwijk, The Netherlands

Bentin S., McCarthy G. and Wood C.C. (1985) Event-related Potentials, Lexical Decision and Semantic Priming. **Electroencephalography and Clinical Neurophysiology. 60: 343-355**

Berger H. (1929) Uber das Elektrenkephalogram des Menschen. **Archiv fur Psychiatrie und Nervenkrankheiten 87:527-570**

Berlin B. and Kay P. (1969) Basic color terms: their universality and evolution. Berkeley, University of California Press.

Besson M., Kutas M. and van Petten C. (1989) Semantic Priming and Word Repetition Effects in Sentence Context: An Event-related Potentials Study. Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX),, Noordwijk, The Netherlands

Bloom P.A. and Fichler I. (1980) Completion norms for 329 sentence contexts. **Memory and Cognition 8:631-642**

Boddy J. (1981) Evoked Potentials and the Dynamics of Language Processing. **Biological Psycholgy 13: 125-140**

Boddy J. (1986) Evoked Potential Correlates of Semantic Priming at Different Stimulus Onset Asynchronies. In: McCallum W.C., Denoth F. and Zappoli R. (Eds.) Cerebral Psychophysiology: Studies in Event-related Potentials (EEG Suppl 38). Elsevier.

Boddy J. and Weinberg H. (1981) Brain Potentials, Perceptual Mechanisms and Semantic Categorisation. **Biological Psychology 12: 43-61**

Braak H. von (1978) Eine ausfuhrliche Beschreibung pigmentachitektonischer Arbeitsverfahren. **Mikroskopie 34:215-221**

Broadbent (1958) Perception and communication. London, Pergamon.

Broadbent D.E. (1967) Word Frequency Effect and Response Bias. **Psychological Review 74: 1-15**

Brown C., Hagoort P. and Swaab T. (1989) Does N400 Reflect Automatic Lexical Processing? Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX),, Noordwijk, The Netherlands

Brown W.S., Lehmann D. and Marsh J.T. (1980) Linguistic meaning related differences in evoked potential topography: English, Swiss-German and imaghined. **Brain and Language 11:340-353**

Brown W.S., Marsh J.T. and Smith J.C. (1976) Evoked potential waveform differences produced by the perception of different meanings of an ambiguous phrase. **Electroencephalography and Clinical Neurophysiology 41:113-123**

Bruyer R. and Janlin D. (1989) Lateral differences in lexical access: Word length vs. stimulus length **Brain and Language 37:258-265**

Bub D.N. (1982) The nature of half-field asymmetry for word processing as a function of hemispheric specialization. Ph.D. Thesis, Rochester University, New York.

Burton M. (1972) Semantic dimensions of occupation names. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Carroll J.D. (1968) A general method for preference mapping of perceptual space. **Bulletin of the Operations Research Society of America 16:282**

Carroll J.D. (1972) Individual differences and multidimensional scaling. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Carroll J.D. and Chang J-J (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. **Psychometrika 35:283-319**

Carroll P. and Slowiaczek M.L. (1986) Contraints on Semantic Priming in Reading: A Fixation Time Analysis. **Memory and Cognition 14: 509-522**

Carroll J.B. and White M.N. (1973) Word frequency and age of acquisition as determiners of picture naming latency. **Quarterly Journal of Experimental Psychology 25:85-95**

Caton R. (1875) The electric currents of the brain. **British Medical Journal 2:278**

Changeux J-P and Danchin A. (1976) Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. **Nature 264:705-711**

Chapman R.M. (1980) Connotative meaning and averaged evoked potentials. In: Evoked Brain Potentials and Behavior, Begleiter H. (Ed.) pp171-190. New York, Plenum.

Chapman R.M., Bragdon H.R., Chapman J.A. and McCrary J.W. (1977) Semantic meaning of words and average evoked potentials. In: Progress in Clinical Neurophysiology vol. 3: Language and Hemispheric Specialization in Man: Cerebral event-related potentials, Desmedt J.E. (Ed.), Basel, Karger.

Chapman R.M., Mcrary J.W., Chapman J.A. and Bragdon H.R. (1978) Brain responses related to semantic meaning. **Brain and Language 5:195-205**

Chiappa K.H. (1983) Evoked potentials in clinical medicine. New York: Raven Press.

Churchland P.S. (1986) Neurophilosophy. Cambridge, MIT Press.

Clark H.H. (1973) The language as fixed effect fallacy: A critique of language statistics in psychological research. **Journal of Verbal Learning and Verbal Behavior 12:335-359**

Cliff N. (1966) Orthogonal rotation to congruence. **Psychometrika 31:33-42**

Cohen J. and Cohen P. (1975) Applied multiple regression/correlation analysis for the behavioral sciences. Potomac, Maryland, Erlbaum.

Coles M.G.H. and Donchin E. (1988) **Behavioral and Brain Sciences**

Coombs C.H. (1964) A theory of data. New York, Wiley.

Cooper R. (1982) Event-related potential methodology. In: Burden Neurological Institute: Brain Evoked Potentials Course. Burden Neurological Institute, Bristol.

Cooper R., Osselton J.W. and Shaw J.C. (1974) EEG technology. London: Butterworth.

Cormack R.M. (1971) A review of classification. **Journal of the Royal Statistical Society, Series A(General) 134:321-367**

Curry S.H. and McCallum W.C. (1982) Cognitive event-related potentials. In: Burden Neurological Institute: Brain Evoked Potentials Course. Burden Neurological Institute, Bristol.

D'Andrade R.G., Quinn N.R., Nerlove B.B. and Romney A.K. (1972) Categories of disease in American-English and Mexican-Spanish. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Darlington R.B. (1968) Multiple regression in psychological research and practice. **Psychological Bulletin 69:161-182**

den Heyer K., Goring A. and Dannenbring G.L. (1985) Semantic priming and word repetition: The two effects are additive. **Journal of Memory and Language 24: 699-716**

Desmedt J.E. (1981) Scalp recorded cerebral event-related potentials in Man as point of entry into the analysis of cognitive processing. In: Schmitt F.O., Worden F.G., Adelman G. and Dennis S.G. (Eds.) Organization of the Cerebral Cortex. MIT Press, Cambridge, Mass.

Desmedt J.E., and Debecker J. (1979) Slow potential shifts and decision P350 interactions in tasks with random sequences of near-threshold clicks and finger stimuli delivered at regular intervals. **Electroencephalography and clinical neurophysiology 47:671-679**

Donald M.W. (1979) Limits on current theories of transient evoked potentials. In: Desmedt J.E. (Ed.) Progress in Clinical Neurophysiology Vol.6:Cognitive Components in Cerebral Event-related Potentials and Selective Attention. Karger, Basel.

Donchin E. (1969) Data analysis techniques in evoked potential research. In: Donchin E. and Lindsley D.B. (Eds.) Average Evoked potentials. NASA sp-191,pp 199-236 (U.S. Goverment printing Office, Washington)

Donchin E., Ritter W. and McCallum W.C. (1978) Cognitive psychophysiology: The endogenous components. In: Callaway E.,Tueting P. and Koslow S.H. (Eds.) Event-related Potentials in Man. New York, Academic Press.

Donchin E., Tueting P., Ritter W., Kutas M. and Heffley E. (1975) On the independence of the CNV and the P300 components of the averaged evoked potential. **Electroencephalography and Clinical Neurophysiology 38:449-461**

Duncan-Johnson C.C. and Donchin E. (1977) On quantifying surprise: The variation of event-related potentials with subjective probability. **Proclamation of the society for psychophysiological research.** San Diego.

Duncan-Johnson C.C. and Kopell B.S. (1981) The Stroop effect: Brain potentials localize the source of interference. **Science 214:938-940**

Dunn G. and Everitt B.S. (1982) An introduction to mathematical taxonomy. Cambridge, Cambridge University Press.

Dyer F.N. (1973) The Stroop phenomenon and its use in the study of perceptual, cognitive and response processes. **Memory and Cognition 1:106-120**

Edelman G.M. (1978) Group selection and phasic reentrant signalling. In: The Mindful Brain: Cortical Organization and the group-selective theory of higher brain function. Mountcastle V.B. and Edelman G.M. (Eds.). Cambridge, MIT Press.

Edgington E.S. (1980) Randomization Tests. Marcel Dekker.

Edwards A.L. (1979) Multiple regression and the analysis of variance and covariance. San Francisco, Freeman.

Etezadi-Amoli J. and McDonald R.P. (1983) A second generation nonlinear factor analysis. **Psychometrika 48:315-342**

Feustel T.C., Shiffrin R.M. and Salasoo A. (1983) Episodic and Lexical Contributions to the Repetition Effect in Word Identification. **Journal of Experimental Psychology: General 112: 309-346**

Fillenbaum and Rapoport (1971) Structures in the subjective lexicon. New York, Academic Press.

Fisher R.A. (1936) The utilization of multiple measurements in taxonomic problems. **Annals of Eugenics 7:179-188**

Fischler I. (1987) Commentary: Progress in brain potential studies of language. In: Johnson R. Jr, Rohrbaugh J.W., and Parasuraman R. Current Trends in Event-related Potential Research (EEG supplementum 40). Elsevier (biomedical division).

Flew A. (1975) Thinking about thinking (or, Do I sincerely want to be right?). Glasgow, Fontana.

Fodor J.A. (1981) Representations: Philosophical essays on the foundations of Cognitive Science. Cambridge, MIT Press.

Fodor J.A. and Pylyshyn Z.W. (1988) Connectionism and cognitive architecture: a critical analysis. **Cognition 28:3-71**

Forster K.I. (1976) Accessing the Mental Lexicon. In: Wales R.J. and Walker E. (Eds.) New Approaches to Language Mechanisms. Amsterdam, North Holland

Forster K.I. and Chambers (1973) Lexical access and naming time. **Journal of Verbal Learning and Verbal Behavior 12:627-635**

Forster K.I. and Davies C. (1984) Repetition Primingand Frequency Attenuation in Lexical Access. **Journal of Experimental Psychology: Learning Memory and Cognition 10: 680-698**

Gaillard A.W.K. (1986) The CNV as an index of response preparation. In: McCallum W.C., Denoth F. and Zappoli R. (Eds.) Cerebral psychophysiology: Studies in Event-related Potentials (EEG suppl. 38) pp196-206. Elsevier.

Gaillard A.W.K. (1988) Problems and paradigms in ERP research. **Biological Psychology 26:91-109**

Galambos R., Benson P., Smith T.S., Schulman-Galambos C. and Osier H. (1975) On hemispheric differences in evoked potentials to speech stimuli. **Electroencephalography and Clinical Neurophysiology 39:279-283**

Gazzaniga M.S. (Ed.) (1984) Handbook of Cognitive Neuroscience. New York, Plenum.

Georgopoulos A.P., Kalaska J.F., Caminiti R. and Massey J.T. (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. **Journal of Neuroscience 2:1527-1537**

Georgopoulos A.P, Kalaska J.F., Crutcher M.D., Caminiti R. and Massey J.T. (1986) The representation of movement direction in the motor cortex: Single cell and population studies. In: Dynamic Aspects of Neocortical Function. Edelman G.M., Gall W.E. and Cowan W.M. (Eds.) pp501-524. New York, Wiley.

Gernsbacher M.A. (1984) Resolving 20 Years of Inconsistent Interactions between Lexical Familiarity and Orthography, Concreteness, and Polysemy. **Journal of Experimental Psychology: General 113: 256-281**

Glucksberg S. and Danks J.H. (1975) Experimental psycholinguistics: an introduction. Hillsdale, NJ, Erlbaum.

Gower J.C. (1971) Statistical methods of comparing different multivariate analyses of the same data. In: Mathematics in the Archaeological and Historical sciences. pp 138-149. Edinburgh.

Grozinger B., Kornhuber H.H., Szirtes J. and Westphal K.T.P. (1980) The Bereitschaftspotential preceding the act of speaking. Also an analysis of artifacts. In: Motivation, Motor and Sensory Processes of the Brain: Electrical Poitentials, Behavior, and Clinical Use: Progress in Brain Research Vol. 54, pp798-804. Amsterdam, Elsevier.

Halgren E. and Smith M.E. (1987) Cognitive evoked potentials as modulatory processes in human memory formation and retrieval. **Human Neurobiology 6:129-139**

Halsey R.M. and Chapanis A. (1954) Chromaticity confusion contours in a complex viewing situation. **Journal of the Optical Society of America 46:442-454**

Hansen J.C. and Hillyard S.A. (1980) Endogenous brain potentials associated with selective auditory attention. **Electroencephalography and Clinical Neurophysiology 49:277-290**

Harter M.R., Aine C. and Schroeder C. (1982) Hemispheric differences in the processing of stimulus location: Effects of selective attention on visual evoked potentials. **Neuropsychologia 20:421-438**

Harter M.R. and Previc F.H. (1978) Size-specific information channels and selective attention: Visual evoked potential and behavioral measures. **Electroencephalography and Clinical Neurophysiology 45:628-640**

Heit G., Smith M.E. and Halgren E. (1988) Neural encoding of individual words and faces by the human hippocampus and amygdala. **Nature 333:773-775**

Henderson L., Wallis J. and Knight D. (1984) Morphemic structure and lexical access. In: Bouma H. and Bouwhuis D. (Eds.) Attention and Performance X. London, Lawrence Erlbaum Associates.

Herning R.I., Speer M. and Jones R.T. (1987) Event-related potentials to spoken equations: Is the N400 really a late N200? In: Current trends in event-related potential research (EEG suppl. 40), Johnson R.Jr., Rohrbaugh J.W. and Parasuraman R. (Eds.). Elsevier.

Hillyard S.A., Hink R.F., Schwent V.L. and Picton T.W. (1973) Electrical signs of selective attention in the brain. **Science 182:177-182**

Hillyard S.A and Munte T.F. (1984) Selective attention to color and location: An analysis with event-related brain potentials. **Perception and Psychophysics 36:185-198**

Hopmann M.R. (1972) Acquisition of basic color terms. Minneapolis, University of Minnesota, Institute of Child Development.

Hotelling H. (1933) Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology 24:417-441,498-520**

Humphreys G.W., Besner D. and Quinlan P.T. (1988) Event Perception and the Word Repetition Effect. **Journal of Experimental Psychology: General. 117: 51-67**

Hunt E. (1979) The validity of assumptions underlying the use of PCA and ANOVA in ERP research. Paper presented to the workshop on analysis of event-related potentials, Vancouver B.C.

Indow T. and Kanazawa K. (1960) Multidimensional mapping of Munsell colors varying in hue, chroma and value. **Journal of Experimental Psychology 59:330-336**

Indow T. and Uchizono (1960) Multidimensional mapping of Munsell colors varying in hue and chroma. **Journal of Experimental Psychology 59:321-329**

Ingvar D.H. (1983) Serial aspects of language and speech related to prefrontal cortical activity: A selective review. **Human neurobiology 2:177-189**

Inhoff A.W. (1984) Two Stages of Word Processing in the Reading of Prose. **Journal of Verbal Learning and Verbal Behavior 23: 612-624**

Jacoby L.L. (1983) Perceptual Enhancement: Persistent Effects of an Experience. **Journal of Experimental Psychology: Learning Memory and Cognition 9: 21-38**

Jasper H.H. (1958) The ten-twenty electrode system. **Electroencephalography and Clinical Neurophysiology 10:371-375**

Jerne N.K. (1967) Antibodies and learning: Selection versus instruction. In: The Neurosciences: A study program. Quarton G.C., Melnechuk T. and Schmidt F.O. pp 200-205. New York, Rockefeller University Press.

Jewett D.L., Romano M.N. and Williston J.S. (1970) Human auditory evoked potentials: Possible brainstem components detected on the scalp. **Science 167:1517-1518**

John E.R., Ruchkin D.S. and Vidal J.J. (1978) Measurement of event-related potentials. In: Event related potentials in Man. Callaway E., Tueting P. and Koslow S.H. (Eds.) pp155-196. New York, Academic Press.

John E.R., Herrington R.N. and Sutton S. (1967) Effects of visual form on the evoked response. Science **155:1439-1442**

Johnson R.Jr. (in press) Auditory and visual P300s in temporal lobectomy patients: Evidence for modality-dependent generators. **Psychophysiology**

Johnson-Laird P.N., Herrmann D.J. and Chaffin R. (1984) Only connections: A critique of semantic networks. **Psychological Bulletin 96:292-315**

Johnstone V.S. and Chesney G.L. (1974) Electrophysiological correlates of meaning. **Science 180:944-946**

Jones G.V. (1985) Deep Dyslexia, Imageability, and Ease of Predication. **Brain and Language 24: 1-19**

Joreskog and Sorbom (1984) Linear Structural Relations Modelling.

Just M.A. and Carpenter P.A. (1980) A Theory of Reading: From Eye Fixations to Comprehension. **Psychological Review 87: 329-354**

Kendall M.G. (1975) Multivariate analysis. Charles Griffin, London.

Kerlinger F.N. and Pedhazur E.J. (1973) Multiple regression in the behavioral sciences. New York, Holt, Rinehart and Winston.

Keselman H.J. and Rogan J.C. (1980) Repeated Measures F tests and Psychophysiological Research: Controlling the Number of False Positives. **Psychophysiology 17: 499-503**

Kinsbourne M. and George J. (1971) The Mechanism of the Word Frequency Effect on Recognition Memory **Journal of Verbal Learning and Verbal Behavior 13: 63-69**

Kintsch W. (1980) Semantic memory: A tutorial. In: Attention and Performance VIII, Nickerson R.S. (Ed.), pp 595-620. Hillsdale, NJ, Erlbaum.

Kirk R.E. (1969) Experimental design: procedures for the behavioral sciences. Belmont, California, Brooks/Cole.

Kliegl R., Olson R.K. and Davidson B.J. (1982) Regression analyses as a tool for studying reading processes: Comment on Just and Carpenter's eye fixation theory. **Memory and Cognition 10:287-296**

Kruskal J.B. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika 29:115-129**

Kucera H. and Francis W.N. (1967) Computational Analysis of Present Day American English. Brown University Press, Providence, RI.

Kutas M. (1987) Commentary: Overview of language and hemispheric specialization. In: Johnson R. Jr, Rohrbaugh J.W., and Parasuraman R. Current Trends in Event-related Potential Research (EEG supplementum 40). Elsevier (biomedical division).

Kutas M. and Hillyard S.A. (1980a) Reading senseless sentences: Brain potentials reflect semantic incongruity. **Science 207:203-205**

Kutas M. and Hillyard S.A. (1980b) Event-related potentials to semantically inappropriate and surprisingly large words. **Biological Psychology 11:99-106**

Kutas M. and Hillyard S.A. (1980c) Reading between the lines: Event-related potentials during natural sentence processing. **Brain and Language 11:354-373**

Kutas M. and Hillyard S.A. (1982) The Lateral Distribution of Event-related Potentials during Sentence Processing. **Neuropsychologia 20: 579-590**

Kutas M. and Hillyard S.A. (1983) Event-related potentials to grammatical errors and semantic anomalies. **Memory and Cognition 11:539-550**

Kutas M. and Hillyard S.A. (1984a) Event-related potentials in cognitive science. In: Gazzaniga M.S. (Ed.) Handbook of Cognitive Neuroscience. Plenum.

Kutas M. and Hillyard S.A. (1984b) Brain potentials during reading reflect word expectancy and semantic association. **Nature 307:161-163**

Kutas M. and Hillyard S.A. (1988) An electrophysiological probe of incidental semantic association. **Journal of Cognitive Neuroscience 1:38-49**

Kutas M. and van Petten C. (in press) Event-related brain potential studies of language. In: Ackles P.K., Jennings J.R. and Coles M.G.H. (Eds.) Advances in Psychophysiology. Greenwich, Conn. JAI Press.

Kutas M., van Petten C. and Besson M. (1988) Event-related Potential Asymmetries during the Reading of Sentences. **Electroencephalography and Clinical Neurophysiology 69: 218-233**

Lindsley D.B. and Wicke J.D. (1974) The electroencephalogram: Autonomous electrical activity in Man and animals. In: Thompson R.F. and Patterson M.M. (Eds.) Bioelectric Recording Techniques, Part B: Electroencephalography and Human Brain Potentials. New York Academic Press.

Lingoes J.C. (1967) An IBM 7090 program for Guttman-Lingoes configurational similarity. **Behavioral Science 12:502-503**

Lovrich D., Novick B. and Vaughan H.G. (1988) Topographic analysis of auditory event-related potentials associated with acoustic and semantic processing. **Electroencephalography and Clinical Neurophysiology 71:40-54**

McCallum W.C. (1986) How many separate processes constitute the CNV. In: McCallum W.C., Denoth F. and Zappoli R. (Eds.) Cerebral psychophysiology: Studies in Event-related Potentials (EEG suppl. 38) pp 192-196. Elsevier.

McCann R.S. and Besner D. (1987) Reading Pseudohomophones: Implications for Models of Pronounciation Assembly and the Locus of Word Frequency Effects in Naming. **Journal of Experimental Psychology: Human Perception and Performance 13: 13-24**

McCann R.S., Besner D. and Davelaar E. (1988) Word Recognition and Identification: Do Word Frequency Effects Reflect Lexical Access? **Journal of Experimental Psychology: Human Perception and Performance 14: 693-706**

McCarthy G. and Donchin E. (1981) A metric for thought: A comparison of P300 latency and reaction time. **Science 211:77-80**

McClelland J.L. (1979) On the time relations of mental processes: An examination of systems of processes in cascade. **Psychological Review 86:287-330**

McClelland J.L. and Elman J.L. (1986) The TRACE model of Speech perception. **Cognitive Psychology 18: 1-86**

McClelland J.L. and Rumelhart D. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. **Psychological Review 88:375-407**

McClelland J.L. and Rumelhart D. (1985) Distributed memory and the representation of general and specific information. **Journal of Experimental Psychology: General 114:159-188**

McDonald R.P. (1962) A general approach to nonlinear factor analysis. **Psychometrika 27:397-415**

McDonald R.P. (1988) Describing the elephant: Structure and function in multivariate data. **Psychometrika 51:513-534**

Maerz A. and Paul M.R. (1930) Dictionary of color. New York, McGraw-Hill.

Manly B.F.J. (1986) Multivariate statistical methods. London, Chapman Hall.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. **Cancer Research 27:209-220**

Marsh J.T. and Brown W.S. (1977) Evoked potential correlates of meaning in the perception of language. In: Progress in clinical neurophysiology Vol 3: Language and Hemispheric Specialization in Man. Desmedt J.E. (Ed.). pp60-72. Basel, Karger.

Mehler J., Morton J. and Jusczyck P. (1984) On reducing language to biology. **Cognitive Neuropsychology 1:83-116**

Meyer D.E., Osman A.M., Irwin D.E. and Yantis S. (1988) Modern mental chronometry. **Biological Psychology 26:3-68**

Meyer D.E., Schvaneveldt R.W. and Ruddy M.G. (1975) Loci of contextual effects on visual word recognition. In: Attention and Performance V, Rabbitt P.M.A. and Dornic S. (Eds.) pp98-118 London, Acacdemic Press.

Mielke P.W. (1978) Classification and appropriate inferences for Mantel's and Valand's nonparametric multivariate analysis technique. **Biometrics 34:272-282**

Miller G.A. and Johnson-Laird P.N. (1976) Language and Perception. Cambridge, Cambridge University Press.

Mocks J. (1988) Decomposing event-related potentials: A new topographic components model. **Biological Psychology 26:199-216**

Molfese D.L. (1983) Event-related potentials and language processes. In: Gaillard A.W.K. and Ritter W. (Eds) Tutorials in ERP research: Endogenous components. North Holland.

Monsell S. (1985) Repetition and the Lexicon. In A.W. Ellis (Ed.) Progress in the Psychology of Language, vol. 2, 147-195 London, Erlbaum.

Monsell S., Doyle M.C. and Haggard P.N. (1989) Effects of Frequency on Visual Word Recognition Tasks: Where are they? **Journal of Experimental Psychology: General in press**

Morris P. (1981) Age of acquisition, imagery, recall, and the limitations of multiple regression analysis. **Memory and Cognition 9:277-282**

Morton J. (1968) A Retest of the Response Bias Explanation of the Word Frequency Effect. **British Journal of Mathematical and Statistical Psychology 21: 21-22**

Morton J. (1969) Interaction of Information in Word Recognition. **Psychological Review 76: 165-178**

Morton J. (1982) Disintegrating the Lexicon: An Information Processing Approach. In: Mahler J., Franck S., Walker E. and Garrett M. (Eds.) Perspectives on Mental Representations. Hillsdale, NJ: Erlbaum

Muller K.E., Otto D.A. and Benignus V.A. (1983) Design and analysis issues and strategies in psychophysiological research. **Psychophysiology 20:212-218**

Naatanen R. (1982) Processing negativity. **Psychological Bulletin 92: 605-640**

Naatanen R. and Gaillard A.W.K. (1983) The orienting reflex and the N2 deflection of the ERP. In: Gaillard A.W.K. and Ritter W. (Eds) Tutorials in ERP research: Endogenous components. North Holland.

Naatanen R., Gaillard A.W.K. and Mantysalo S. (1978) Early selective attention effect on evoked potential reinterpreted. **Acta Psychologica 42:313-329**

Naatanen R. and Picton T. (1986) N2 and automatic versus controlled processes. In: McCallum W.C., Zappoli R. and Denoth F. (Eds.) Cerebral Psychophysiology: Studies in event-related potentials (EEG supplementum 38). Elsevier (Biomedical division).

Nagy M.E. and Rugg M.D. (1989) Modulation of Event-related Potentials by Word Repetition: Effects of Inter Item Lag. **Psychophysiology in press**

Nerlove S.B. and Romney A.K. (1972) Introduction. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Neville H. (1980) Event-related potentials in neuropsychological studes of language. **Brain and Language 11:300-318**

Norris D. (1984) The Effects of Frequency, Repetition and Stimulus Quality in Visual Word Recognition. **Quarterly Journal of Experimental Psychology 36A: 507-518**

Okita T., Wijers A.A., Mulder G. and Mulder L.J.M. (1985) Visual search and spatial attention: ERPs in focussed and divided attention conditions. **Acta Psychologica 60:263-292**

Oliphant G.W. (1983) Repetition and Recency Effects in Word Recognition. **Australian Journal of Psycholology 35: 393-403**

Osgood C.E. (1971) Exploration in semantic space: A personal diary. **Journal of Social Issues 27:5-64**

Osgood C.E., May W.H. and Miron M.S. (1975) Cross-cultural universals of affective meaning. Urbana, Illinois, University of Illinois Press.

Osgood C.E., Suci G.J. and Tannenbaum P. (1957) The measurement of meaning. University of Illinois Press, Urbana.

Osterhout L. and Holcomb P. (1989) Event-related potentials elicited by garden path sentences. Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX),, Noordwijk, The Netherlands

Paivio A. Yuille J. and Madigan S. (1968) Concreteness, Imagery and Meaningfulness Values for 925 Nouns. **Journal of Experimental Psychology Monograph 76:(1, Pt. 2)**

Pearson K. (1901) On lines and planes of closest fit to a system of points in space. **Philosophical Magazine 2:557-572**

Petersen S.E., Fox P.T., Posner M.I., Mintun M. and Raichle M.E. (1988) Positron emission tomographic studies of the processing of single words. **Journal of Cognitive Neuroscience 1:153-170**

Picton T.W. and Stuss D.T. (1984) Event-related potentials in the study of speech and language: A critical review. In: Biological Perspectives on Language. Caplan D., Lecours A. and Smith A. (Eds.) pp303-360. Cambridge, Mass, MIT Press.

Polich J. and Donchin E. (1988) P300 and the Word Frequency Effect. **Electroencephalography and Clinical Neurophysiology 70: 33-45**

Polich J., Vanasse L. and Donchin E. (1981) Category expectancy and the N200. **Psychophysiology 18:142ff**

Pritchard W.S. (1981) The psychophysiology of P300. **Psychological Bulletin 89:506-540**

Purves S.J., Low M.D. and Baker M. (1979) Evoked potential correlates of stimulus information content. In: Desmedt J.E. (Ed.) Progress in Clinical Neurophysiology Vol.6:Cognitive Components in Cerebral Event-related Potentials and Selective Attention. pp172-181. Karger, Basel.

Pylyshyn Z.W. (1984) Computation and cognition. MIT Press. Cambridge, Mass.

Rapoport A. and Fillenbaum S. (1972) An experimental study of semantic structures. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Rayner K. and Pollatsek A. (1981) Eye Movement Control During Reading: Evidence for Direct Control. **Quarterly Journal of Experimental Psychology 33A: 351-374**

Regan D. (1972) Evoked potentials in psychology, sensory physiology, and clinical medicine. Chapman and Hall, London.

Renault B. and Lesevre N. (1979) Topographical study of the emitted potential obtained after the omission of an expected visual stimulus. In: Otto D. (Ed.) Multidisciplinary perspectives in event-related potential research. Washington, U.S. Government printing office.

Ritter W. (in press) In Proceedings of the Ninth International Congress on Event-related potentials of the brain, Noorwijk, The Netherlands.

Ritter W., Ford J.M., Gaillard A.W.K., Harter M.R., Kutas M., Naatanen R., Polich J., Renault B. and Rohrbaugh J. (1984) Cognition and event-related potentials: I. The relation of negative potentials and cognitive processes. In: Karrer R., Cohen J. and Tueting P. (Eds.) Brain and Information: Event Related Potentials. New York, N.Y. Academy of sciences.

Richards L.G. and Hempstead J.D. (1973) Auditory pretraining as a determinant of visual thresholds for pseudowords. **American Journal of Psychology 86:325-329**

Ritter W., Simson R. and Vaughan H.G. (1983) Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. **Psychophysiology 20:168-179**

Ritter W., Simson R., Vaughan H.G. and Friedman D.A. (1979) A brain event related to the making of a sensory discrimination. **Science 203:1358-1361**

Ritter W., Simson R., Vaughan H.G. and Macht M. (1982) Manipulation of evnt-related potential manifestations of information processing stages. **Science 218:909-911**

Roemer R. and Teyler T. (1977) Auditory evoked potential asymmetries related to word meaning. In Language and Hemispheric Specialization in Man. **Progress in clinical Neurophysiology 3:48-59**

Rosenberg S. and Sedlak A. (1972) Structural representations of perceived personality trait relationships. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Rosler F. and Manzey D. (1981) Principal components and varimax rotated components in event-related potential research; some remarks on their interpretation. **Biological Psychology 13:3-26**

Rubenstein H., Garfield L. and Millikan J.A. (1970) Homographic Entries in the Internal Lexicon. **Journal of Verbal Learning and Verbal Behavior 9: 487-494**

Rubin D.C. (1980) 51 Properties of 125 Words: A Unit Analysis of Verbal Behavior. **Journal of Verbal Learning and Verbal Behavior 19: 736-755**

Rugg M.D. (1983a) Further study of the electrophysiological correlates of lexical decision. **Brain and Language 19:142-153**

Rugg M.D. (1983b) The relationship between evoked potentials and lateral asymmetries of processing. In: Tutorials in ERP research: Endogenous components. Gaillard A.W.K. and Ritter W. (Eds.). Amsterdam, North Holland.

Rugg M.D. (1984) Event-related potentials and the phonological processing of words and nonwords **Neuropsychologia 22:435-443**

Rugg M.D. (1985) The Effects of Semantic Priming and Word Repetition on Event-related Potentials. **Psychophysiology 22: 642-647**

Rugg M.D. (1987) Dissociation of Semantic Priming, Word and Nonword Repetition Effects by Event-related Potentials. **Quarterly Journal of Experimental Psychology 39A: 123-148**

Rugg M.D. (in press) Event-related Brain Potentials Dissociate Repetition Effects with High and Low Frequency Words. **Memory and Cognition**

Rugg M.D. (in press) Event-related potentials in Clinical Neurophysiology. In: Crawford J.R., McKinley W.A. and Parker D.M. (Eds.) Practice and Principles of Neuropsychological Assessment. London: Taylor and Francis.

Rugg M.D. and Barrett S.E. (1987) Event-related potentials and interaction between orthographic and phonological information in a rhyme-judgement task. **Brain and Language 32:336-361**

Rugg M.D., Furda J. and Lorist M. (1988) The Effects of Task on the Modulation of Event-related Potentials by Word Repetition. **Psychophysiology 25: 55-63**

Rugg M.D., Kok A., Barrett G., and Fischler I. (1986) ERPs associated with language and hemispheric specialisation. A review. In: McCallum W.C., Zappoli R. and Denoth F. (Eds.) Cerebral Psychophysiology: Studies in event-related potentials (EEG supplementum 38). Elsevier (Biomedical division).

Rugg M.D., Milner A.D, Lines C.R and Phalp R. (1987) Modulation of visual event-related potentials by spatial and non-spatial visual selective attention **Neuropsychologia 25:85-96**

Rugg M.D. and Nagy M.E. (1987) Lexical Contribution to Nonword Repetition Effects: Evidence from Event-related Potentials. **Memory and Cognition 15: 473-481**

Scarborough D.L, Cortese C. and Scarborough H.S. (1977) Frequency and Repetition Effects in Lexical Memory. **Journal of Experimental Psychology: Human Perception and Performance 3: 1-17**

Schonemann P. (1966) A generalized solution of the orthogonal Procrustes problem. **Psychometrika 31:1-10**

Schonemann P. and Carroll R.M. (1970) Fitting one matrix to another under choice of a similarity transformation and a rigid motion. **Psychometrika 35:245-255**

Searle J.R. (1980) Minds, brains and programs. **Behavioral and Brain Sciences 3:417-457**

Searle J.R. (1983) Intentionality: An essay in the philosophy of mind. Cambridge, Cambridge University Press.

Seidenberg M.S., Waters G.S., Barnes M.A. and Tannenhaus M.K. (1984) When does irregular spelling or pronunciation influence word recognition. **Journal of Verbal Learning and Verbal Behavior 23:383-404**

Seymour P.H.K. (1977) Conceptual encoding and the locus of the Stroop effect. **Quarterly Journal of Experimental Psychology**

Seymour P.H.K. and May G.P. (1981) Locus of format distortion effects in word recognition. Paper presented to the Experimental Psychology Society, Oxford.

Shepard R.N. (1962) The analysis of proximities: Multidimensional scaling with an unknown distance function. I and II. **Psychometrika 27:125-140, 219-246**

Shepard R.N. (1972) A taxonomy of some principal types of data and of multidimensional methods for their analysis. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Shepard R.N. (1980) Multidimensional scaling, tree-fitting and clustering. **Science 210:390-398**

Shepard R.N. and Carroll J.D. (1966) Parametric representation of nonlinear data structures. In: International symposium on multivariate analysis, Dayton, Ohio, Krishnaiah P.R. (Ed.). New York, Academic Press.

Shepard R.N., Nerlove S.B. and Romney A.K. (1972) Multidimensional scaling: Theory and application in the behavioral sciences. New York, Seminar Press.

Smith M.E. and Halgren E. (1987) Event-related Potentials during Lexical Decision: Effects of Repetition, Word Frequency, Pronounceability and Concreteness. In: Current Trends in Event-related Potential Research (EEG suppl. 40) Johnson R., Rohrbaugh, J.W. and Parasuraman R. (Eds.)

Stanners R.F., Forbach G.B. and Headley D.B. (1971) Decision and search processes in word-nonword classification. **Journal of Experimental Psychology 90:45-50**

Stanners R.F., Jastrzembski J.E. and Westbrook A. Frequency and visual quality in a word-nonword classification task. **Journal of Verbal Learning and Verbal Behavior 14:259-264**

Starr A., Sohmer H. and Celesia G.G. (1978) Some applications of evoked potentials to patients with neurological and sensory impairment. In: Event related potentials in Man. Callaway E., Tueting P. and Koslow S.H. (Eds.) pp155-196. New York, Academic Press.

Stenson H.H. and Knoll R.L. (1969) Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. **Psychological Bulletin 71:122-126**

Stephenson W.A. and Gibbs F.A. (1951) A balanced non-cephalic reference electrode. **Electroencephalography and Clinical Neurophysiology 3:237-240**

Sternberg S. (1969) The Discovery of Processing Stages: Extensions of Donders' Method. In: Koster W.G.(Ed.) Attention and Performance II pp58-87. Amsterdam: North Holland.

Stirling (1979) Stroop interference: An input and an output phenomenon. **Quarterly Journal of Experimental Psychology**

Stroop J.R. (1935) Studies of interference in serial verbal reactions. **Journal of Experimental Psychology 18:643ff**

Stuss D.T., Picton T.W. and Cerri A.M. (1988) Electrophysiological Manifestations of Typicality Judgement. **Brain and Language 33: 260-272**

Stuss D.T., Sarazin F.F., Leech E.E. and Picton T.W. (1983) Event-related potentials during naming and mental rotation. **Electroencephalography and Clinical Neurophysiology 56:133-146**

Sutton S., Tueting P., Zubin J. and John E.R. (1967) Information delivery and the sensory evoked potential. **Science 155:1436-1439**

Szirtes J. and Vaughan H.G. (1977) Characteristics of cranial and facial potentials associated with speech production. In Language and Hemispheric Specialization in Man. **Progress in Clinical Neurophysiology 3:112-126**

Thatcher R.W. and John E.R. (1976) Functional Neuroscience. Lawrence Erlbaum, New Jersey.

Thorndike E.L. Johnson and Lorge I. (1944) The Teacher's word book of 30,000 words. New York, Teachers College Press.

Treisman M. (1971) On the Word Frequency Effect: Comment on the papers by J. Catlin and L.H. Nakatani **Psychological Review 78: 420-425**

Tominaga D., Kuda K., Yogi H., Miyari T. and Tomori A. (1989) Changes of ERP(P3) and RT during the repeated Stroop colour-word interference task. Abstract and poster presentation at the ninth international congress on event-related potentials of the brain (EPIC IX),, Noordwijk, The Netherlands

Tukey J.W. (1978) A data analyst's comments on a variety of points and issues. In: Event related potentials in Man. Callaway E., Tueting P. and Koslow S.H. (Eds.) pp155-196. New York, Academic Press.

Van Petten C. and Kutas M. (in press) Interactions between sentence context and word frequency in event-related brain potentials. **Memory and Cognition**

Verleger R.(1988) Event-related potentials and memory: A critique of the context updating hypothesis and an alternative interpretation of P3. **Behavioral and Brain Sciences**

Warren R.E. (1972) Stimulus encoding and memory. **Journal of Experimental Psychology 94:90-100**

Wastell D.G. (1981) PCA and varimax rotation; some comments on Rosler and Manzey. **Biological Psychology 13:27-29**

Watkins M.J. (1984) Models as toothbrushes. **Behavioral and Brain Sciences 7:86**

Whaley C.P. (1978) Word-Nonword Classification Time. **Journal of Verbal Learning and Verbal Behavior 17: 143-154**

Wish M., Deutsch M. and Biener L. (1972) Differences in perceived similarity of nations. In: Multidimensional Scaling: Theory and applications in the behavioral sciences. Romney A.K., Shepard R.N. and Nerlove S.B. (Eds.). New York, Seminar Press.

Wood C.C. (1987) Generators of Event-related potentials. In: Halliday A.M., Butler S.R. and Paul R. (Eds.) A Textbook of Clinical Neurophysiology. Wiley.

Wood C.C. and McCarthy G. (1984) Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. **Electroencephalography and Clinical Neurophysiology 59:249-260**

Wood C.C., McCarthy G., Squires N.K., Vaughan H.G., Woods D.L. and McCallum W.C. (1984) Anatomical and physiological substrates of event-related potentials: Two case studies. **Annals of the New York Academy of Sciences 425:681-721**

Yamane S., Kaji S. and Kawano K. (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? **Experimental Brain Research 73:209-214**

Young J.Z. (1973) Memory as a selective process. In: Australian Academy of Science Report: Symposium on Biological Memory. pp25-45. Canberra, Australian Academy of Science.

Young A.W. and Ellis A.W. (1985) Different methods of lexical access for words presented in the left and right visual hemifields. **Brain and Language 24:326-358**

Young M.P. and Routh D.A. (unpublished observations) Inteference in a Stroop-type task as a function of the semantic distance between hue and distractor. (Undergraduate dissertation available from author.)

Young F.W., Takane Y. and Lewyckyj R. (1978) ALSCAL: A nonmetric multidimensional scaling program with several differences options. **Behavior Research Methods and Instrumentation 10:451-453**

Appendix 1: Control language for the data reduction and data analysis procedures

1/. Control language for the data reduction procedures for SPSS-X

```
TITLE ERP>PROX>INDSCAL EACH S>PROX>INDSCAL GLOBAL
SUBTITLE  ALL S experiment 2 400 to 800 msec
DATA LIST / RED 1-5 YELLOW 11-15 GREEN 16-20 BLUE 21-25
        PURPLE 26-30 GREY 36-40 BLACK 41-45 WHITE 46-50
        CHANNEL 51-55 SUB 56-60
BEGIN DATA

                The above are header and data format
                instructions. The ERP data is inserted
                in this part of the input file between
                'BEGIN DATA' and 'END DATA' statements.


END DATA
SPLIT FILE BY SUB CHANNEL               mark data by subject and
                                        by electrode channel
PROXIMITIES RED TO WHITE
 / VIEW = VARIABLE                      call the profile
 / PRINT= NONE                          comparison procedure and
 / MATRIX OUT(*)                        outfile the proximity
                                        data between conditions
                                        for each channel
SPLIT FILE BY SUB
ALSCAL VARIABLES = RED TO WHITEcall the MDS routine to
 / MODEL = INDSCAL                      derive the within subject
 / CRITERIA = DIMENS(3)                 variance model for each
 / OUTFILE = STRUCS                     subject. Then outfile the
                                        model for each subject
GET FILE = STRUCS / MAP
 / KEEP = MATNUM_ DIM1 DIM2 DIM3 SUB          get the relevant
 / MAP                                        parts of the data
SELECT IF (MATNUM_ EQ 0)                      file and check it
SPLIT FILE BY SUB
PRINT / DIM1 DIM2 DIM3 (3(F12.8)) SUB (DOLLAR8,1X)

PROXIMITIES DIM1 DIM2 DIM3                    call the profile
 / VIEW = CASE                          comparison routine to
 / PRINT = NONE                         derive a single within
 / MATRIX OUT(*)                        subject proximity matrix
                                        and outfile it
SPLIT FILE OFF
```

```
ALSCAL VARIABLES = CASE1 TO CASE8        call the MDS routine
 / MODEL = INDSCAL                       to derive the global
 / CRITERIA = DIMENS(3)                  variance model from the
 / OUTFILE = TEMP                        within subject proximity
GET FILE = TEMP                          data and write it to a
 / KEEP = MATNUM_ DIM1 DIM2 DIM3         file to be used by the
SELECT IF ( MATNUM_ EQ 0)                PROCRUSTES data analysis
WRITE OUTFILE = COORDS                   procedures
 / DIM1 DIM2 DIM3 (3(F7.4))
EXECUTE
```

2/. Control language for the data analysis procedures written in the GENSTAT statistical programming language

```
'REFE/NUNN=500' DGMr124                                  title

'SCAL' MSQ,DIM,WEX,WEY,WEZ,WJX,WJY,WJZ,CRIT,COUNT,PROP
  :  VI,VJ,VZ,VY,VX,INC,NO,ONE                           declare variables

'VALU' DIM=3
  :  WEX=0.1566                                          assign values
  :  WEY=0.1670
  :  WEZ=0.1361

'MATR' E,J $ 8,DIM                                       declare matrices
  :  TEMPE,TEMPJ $ 3,8
  :  R $ 8,1
  :  XOUT,YOUT $ 8,3
  :  LOADS $ 3,3
'VARI' EX,EY,EZ,JX,JY,JZ $ 8
'READ' EX,EY,EZ $ F,7,7,7,/                              read model into ERP
'RUN'                                                    data variables
```

The global ERP variance model for latency window which is the subject of analysis is inserted at this point between 'RUN' and 'EOD'

```
'EOD'
```

```
’READ’ JX,JY,JZ $ F,7,7,7,/                  read in the first of the
’RUN’                                        linguistic factor models
 7.2625 0.0000 0.0000                        against which the ERP model
 1.8457 0.0000 0.0000                        is to be rotated
 5.8933 0.0000 0.0000
 8.0179 0.0000 0.0000
 6.3200 0.0000 0.0000
 4.7739 0.0000 0.0000
 2.3833 0.0000 0.0000
 9.3683 0.0000 0.0000
’EOD’

’CALC’ EX=EX*WEX                             stretch the ERP model  according
 :   EY=EY*WEY                               to the ’overall importance of
 :   EZ=EZ*WEZ                               each dimension’

’EQUA’ TEMPE=EX,EY,EZ                        write the ERP model into
’CALC’ E=TRANS(TEMPE)                        an appropriate matrix

’EQUA’ TEMPJ=JX,JY,JZ                        do the same for the factor model
’CALC’ J=TRANS(TEMPJ)                        (only matters if the model has
                                             more than one dimension)

’ROTATE/PRIN=S,SCALE=Y’ E;J;MFIT=CRIT                call the
                                                     PROCRUSTES routine
’PRINT’ CRIT                                 to compare the 2 configurations

’CALC’ VJ=RANDU(5)                           initialize random number
’CALC’ COUNT=0                               generator for the randomization
’VALU’ NO=8                                  test for this linguistic factor
 :   ONE=1                                   model
’RUN’

’FOR’  II = 1...400                          400 iterations in randomization
                                             test

’FOR’  JJ=1...8
’CALC’ VI=RANDU(0)
 :   VI=VI*NO
 :   VI=VI+ONE
 :   VI=INTPT(VI)
 :   VX=ELEM(JX;VI)
 :   VY=ELEM(JY;VI)                          permute the rows of the
 :   VZ=ELEM(JZ;VI)                          linguistic factor model
’CALC’  ELEM(JX;VI)=ELEM(JX;JJ)              at random every iteration
 :    ELEM(JY;VI)=ELEM(JY;JJ)
 :    ELEM(JZ;VI)=ELEM(JZ;JJ)
 :    ELEM(JX;JJ)=VX
 :    ELEM(JY;JJ)=VY
 :    ELEM(JZ;JJ)=VZ

’REPE’                                       continue

’EQUA’  TEMPJ=JX,JY,JZ
’CALC’  J=TRANS(TEMPJ)
```

```
'ROTATE/SCALE=Y' E;J;MFIT=MSQ             compare the 2 models by
                                          PROCRUSTES rotation every
                                          iteration

'CALC'  INC=MSQ .LT. CRIT                 increment the counter if
 :    COUNT=COUNT+INC                     the permuted data fits
                                          the ERP model better than
                                          the 'real' model
'REPE'

'CALC'  PROP=COUNT/II                     compute the probability
                                          that the linguistic
                                          factor model fitted the
                                          global ERP variance model
                                          by chance


'' ***************************MEAN DIGRAM FREQUENCY****** ''
'PRINT' COUNT
 :    II
 :    PROP
'RUN'                                     go on to the next model
'READ' JX,JY,JZ $ F,7,7,7,/
'RUN'
 3.7710 0.0000 0.0000
 1.0930 0.0000 0.0000
 1.1.....
       .
       .
       .
       .


'CLOSE'
'STOP'
```

Appendix 2: Numerical values of the lingustic factor models for the colour names

Mean digram frequency

7.2625	0.0000	0.0000
1.8457	0.0000	0.0000
5.8933	0.0000	0.0000
8.0179	0.0000	0.0000
6.3200	0.0000	0.0000
4.7739	0.0000	0.0000
2.3833	0.0000	0.0000
9.3683	0.0000	0.0000

First letter frequency

3.7710	0.0000	0.0000
1.0930	0.0000	0.0000
1.1010	0.0000	0.0000
1.0790	0.0000	0.0000
1.2830	0.0000	0.0000
1.1010	0.0000	0.0000
1.0790	0.0000	0.0000
1.2090	0.0000	0.0000

Last letter frequency

2.0520	0.0000	0.0000
1.2090	0.0000	0.0000
4.3800	0.0000	0.0000
8.1810	0.0000	0.0000
8.1810	0.0000	0.0000
1.0930	0.0000	0.0000
0.2960	0.0000	0.0000
8.1810	0.0000	0.0000

Mean letter frequency

4.6680	0.0000	0.0000
3.4312	0.0000	0.0000
5.1228	0.0000	0.0000
3.3720	0.0000	0.0000
3.1243	0.0000	0.0000
3.5360	0.0000	0.0000
2.1326	0.0000	0.0000
4.7694	0.0000	0.0000

Kucera-Francis 1

```
1.9700 0.0000 0.0000
0.5500 0.0000 0.0000
1.1600 0.0000 0.0000
1.4300 0.0000 0.0000
0.1300 0.0000 0.0000
0.1200 0.0000 0.0000
2.0300 0.0000 0.0000
3.6500 0.0000 0.0000
```

Kucera-Francis 2

```
1.5000 0.0000 0.0000
1.2000 0.0000 0.0000
1.3000 0.0000 0.0000
1.4000 0.0000 0.0000
0.6000 0.0000 0.0000
0.6000 0.0000 0.0000
1.4000 0.0000 0.0000
1.4000 0.0000 0.0000
```

Kucera-Francis 3

```
1.0100 0.0000 0.0000
0.4100 0.0000 0.0000
0.7700 0.0000 0.0000
0.8500 0.0000 0.0000
0.0900 0.0000 0.0000
0.0800 0.0000 0.0000
1.0000 0.0000 0.0000
1.5600 0.0000 0.0000
```

Thorndike-Lorge

```
1.0360 0.0000 0.0000
0.5970 0.0000 0.0000
1.0250 0.0000 0.0000
0.9890 0.0000 0.0000
0.1320 0.0000 0.0000
0.0280 0.0000 0.0000
1.0830 0.0000 0.0000
2.6630 0.0000 0.0000
```

Number of letters

```
 3.0000 0.0000 0.0000
 6.0000 0.0000 0.0000
 5.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 6.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 5.0000 0.0000 0.0000
 5.0000 0.0000 0.0000
```

Number of phonemes

```
 3.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 3.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 3.0000 0.0000 0.0000
 4.0000 0.0000 0.0000
 3.0000 0.0000 0.0000
```

Number of syllables

```
 1.0000 0.0000 0.0000
 2.0000 0.0000 0.0000
 1.0000 0.0000 0.0000
 1.0000 0.0000 0.0000
 2.0000 0.0000 0.0000
 1.0000 0.0000 0.0000
 1.0000 0.0000 0.0000
 1.0000 0.0000 0.0000
```

Orthographic similarity

```
-1.1564-1.0499 0.3274
 0.9623 0.0481-1.6222
-1.1611-0.7490-0.1994
 0.8165 0.4349 1.3329
-0.1719 1.7028-1.0491
-1.2709-0.2315 0.0290
 0.7522 1.1832 1.4978
 1.2294-1.3387-0.3163
```

Phonological similarity

```
 1.1409-0.1798-1.3488
-0.0033-1.3946 1.0420
 1.0040 0.7213 1.1334
-1.4560 0.0252 0.1056
-0.0930-1.5868-0.9686
 1.0683 0.5825 1.0721
-1.5520 0.2210 0.2487
-0.1089 1.6111-1.2845
```

Semantic similarity

```
-0.4340-2.0125 0.4539 1
 1.3506 0.0392 0.7984 1
 0.2661 1.2045 1.4153 1
-0.7622 1.0330 0.7160 1
-1.2030-0.8682 0.2330 1
 0.3746 0.4492-1.4238 1
-1.1498 0.5584-1.3749 1
 1.5577-0.4036-0.8180 1
```

Appendix 3: Numerical values for the linguistic factor models for the furniture terms

Mean digram frequency

0.3900	0.0000	0.0000
0.2914	0.0000	0.0000
0.4148	0.0000	0.0000
1.1002	0.0000	0.0000
0.4595	0.0000	0.0000
0.6750	0.0000	0.0000
0.6139	0.0000	0.0000
0.3542	0.0000	0.0000
0.2816	0.0000	0.0000
0.3368	0.0000	0.0000

First letter frequency

3.9760	0.0000	0.0000
1.7380	0.0000	0.0000
1.7380	0.0000	0.0000
6.4180	0.0000	0.0000
3.9760	0.0000	0.0000
1.0790	0.0000	0.0000
1.2090	0.0000	0.0000
3.9760	0.0000	0.0000
2.0520	0.0000	0.0000
1.7380	0.0000	0.0000

Last letter frequency

4.8840	0.0000	0.0000
2.0520	0.0000	0.0000
3.7710	0.0000	0.0000
8.1810	0.0000	0.0000
2.6660	0.0000	0.0000
2.0520	0.0000	0.0000
8.1810	0.0000	0.0000
2.0520	0.0000	0.0000
0.2960	0.0000	0.0000
3.5020	0.0000	0.0000

Mean letter frequency

```
3.7650 0.0000 0.0000
2.6426 0.0000 0.0000
3.6864 0.0000 0.0000
4.6456 0.0000 0.0000
4.5208 0.0000 0.0000
3.7707 0.0000 0.0000
3.7149 0.0000 0.0000
3.9227 0.0000 0.0000
3.6263 0.0000 0.0000
2.6624 0.0000 0.0000
```

Kucera-Francis 1

```
0.0600 0.0000 0.0000
0.0200 0.0000 0.0000
0.6600 0.0000 0.0000
1.9800 0.0000 0.0000
0.0800 0.0000 0.0000
1.2700 0.0000 0.0000
0.0800 0.0000 0.0000
0.0100 0.0000 0.0000
0.6500 0.0000 0.0000
0.1200 0.0000 0.0000
```

Kucera-Francis 2

```
0.0300 0.0000 0.0000
0.0200 0.0000 0.0000
0.1300 0.0000 0.0000
0.1400 0.0000 0.0000
0.0300 0.0000 0.0000
0.1300 0.0000 0.0000
0.0400 0.0000 0.0000
0.0100 0.0000 0.0000
0.1000 0.0000 0.0000
0.0600 0.0000 0.0000
```

Kucera-Francis 3

```
0.0300 0.0000 0.0000
0.0200 0.0000 0.0000
0.4800 0.0000 0.0000
1.0600 0.0000 0.0000
0.0500 0.0000 0.0000
0.7000 0.0000 0.0000
0.0600 0.0000 0.0000
0.0100 0.0000 0.0000
0.4400 0.0000 0.0000
0.0900 0.0000 0.0000
```

Number of letters

```
4.0000 0.0000 0.0000
8.0000 0.0000 0.0000
5.0000 0.0000 0.0000
5.0000 0.0000 0.0000
5.0000 0.0000 0.0000
3.0000 0.0000 0.0000
8.0000 0.0000 0.0000
9.0000 0.0000 0.0000
4.0000 0.0000 0.0000
5.0000 0.0000 0.0000
```

Number of phonemes

```
4.0000 0.0000 0.0000
6.0000 0.0000 0.0000
3.0000 0.0000 0.0000
4.0000 0.0000 0.0000
4.0000 0.0000 0.0000
3.0000 0.0000 0.0000
5.0000 0.0000 0.0000
6.0000 0.0000 0.0000
4.0000 0.0000 0.0000
3.0000 0.0000 0.0000
```

Number of syllables

```
2.0000 0.0000 0.0000
2.0000 0.0000 0.0000
1.0000 0.0000 0.0000
2.0000 0.0000 0.0000
1.0000 0.0000 0.0000
1.0000 0.0000 0.0000
2.0000 0.0000 0.0000
2.0000 0.0000 0.0000
1.0000 0.0000 0.0000
1.0000 0.0000 0.0000
```

Orthographic similarity

0.6410-0.2917 1.5870
-1.4184 0.1082-0.6946
0.6662-0.2345-1.7076
0.6804-1.4858 0.4435
0.5936-1.2820 0.8703
0.6344 1.8034 0.3903
-1.6221 0.2396 0.0259
-1.5285-0.1017 0.3778
0.8037 1.5976 0.3532
0.5497-0.3531-1.6458

Phonological similarity

-1.0286 1.1650 0.8677
1.3982 0.5612-0.4430
-0.8948 0.0169-1.5659
-0.0601-0.7440 1.7233
-1.0342 0.8957 1.1244
-0.1184-1.7787 0.1317
1.5531 0.5481-0.1124
1.3586 0.6957 0.0475
-0.3112-1.7043-0.2702
-0.8626 0.3445-1.5031

Semantic similarity

0.9237 0.7131-0.9897
-1.6278 0.0448-0.2715
0.7884 1.0899 0.7494
0.2191-1.5678 1.0124
0.6389 1.1480 1.2651
0.6919-0.0034-2.0252
-1.6489 0.3827-0.2893
-1.1510-1.0169 0.3041
0.2528-1.6266 0.9745
0.9127 0.8362-0.7298

Appendix 4: Word list for experiment 6 (Chapter 6)

List One

CLAY	CONCERN
WALK	WEATHER
JAZZ	JANUARY
POET	POPULAR
FILM	FAILURE
PARK	PARENTS
SIGN	SUPPOSE
LORD	LIBRARY
THIN	TALKING
BASE	BALANCE
PASS	PATIENT
PAIN	PRODUCT
ROSE	RUSSIAN
BANK	BENEFIT
TEAM	TEACHER
TRIP	TRAFFIC
TERM	TYPICAL
FAST	FACULTY
DUST	DOLLARS
CAMP	COMPLEX
MINE	MEASURE
KING	KITCHEN
PAGE	PASSING
HUNG	HIGHEST
BABY	BILLION
TASK	TUESDAY
COOL	CONTACT
ITEM	IMAGINE
DROP	DESIRED
PALE	PAYMENT
WAGE	WELFARE
LAKE	LARGEST
SOIL	SUGGEST
ENDS	EXTREME
BAND	CULTURE
COOK	CHINESE
TONE	OCTOBER
TINY	TRAINED
NINE	NOTICED
REAR	REASONS

List Two

MARK	MANAGER
POST	PRIMARY
MOON	MUSICAL
VOTE	VARIETY
GREW	GERMANY
ROCK	REGULAR
SEND	SESSION
EDGE	ECONOMY
RICH	REALITY
RISK	RECEIVE
RULE	REALIZE
CELL	COMMAND
SITE	SOMEHOW
VAST	VILLAGE
SONG	SHELTER
LOSE	LIBERAL
GAIN	GREATLY
BOAT	BEDROOM
PLUS	PERFECT
FOOT	FASHION
DRAW	DISEASE
FLOW	FIFTEEN
DESK	DEVOTED
SHOP	STORIES
FUND	FORMULA
TEXT	THEATRE
PURE	PORTION
AGES	ACHIEVE
BUSY	BRITAIN
BEAR	ABILITY
PICK	PERCENT
SEAT	SILENCE
SICK	SUPREME
EVIL	EVIDENT
CARS	CORRECT
TREE	TELLING
TALL	TRAGEDY
TILL	TREATED
RUNS	RAPIDLY
SAFE	SECTION

Nonword targets

IMOTIAL
CORTUNT
MINRION
MAPINUM
INAPIAN
ERAMENT
SHIPHAS
NALK
RAFE
JOLL
SNUX
WOUF
CEND
RAOR
EXPROIN
VARSOON
WESCIME
ABIXATY
ESPICLY
OSRABER
ARCAMPT
RADI
SEAC
YOND
HOMI
DEET
HULD
VOLL
UNAPUAL
ELAMERT
SHURCHO
CLIPTER
CONSICT
AMIPERN
CRATSAS
POIR
CIAT
TEVE
TEFE
FIAR
JIRP
ESTALLY
OSTOCER
SPINIAN
ATTOCLE
ALSENTO
ESRUNSA
MIPINAL
FAIG
FRAT
FATY
SPUV
JAGE
REES
CANLUCT
UMIFERM
ARTOMPT
OMOTIAK
COLTINE
MANRIAC
FOPE
WOGE
HACY
TURS
VOLD
KIVS
SIAR
ACSONCE
EXPURSE
CRATHES
AXPRAIL
VORSOAN
WASLINE
REAT
LAON
HUIN
PUST
GURT
BAPE
GOLL

Appendix 5: Word list for experiment 7 (Chapter 7)

High frequency words

FIRST	MAN	YEARS	PEOPLE	HOUSE	LITTLE
STATE	USE	WORLD	WORK	DAY	PLACE
HIGH	WAR	AWAY	REASON	SOUTH	LOVE
TRUE	AGE	WOMAN	CONTROL	SIX	MUSIC
YOUNG	ART	PLAN	LEAVE	MILLION	MODERN
ALONE	AMERICA	BASIC	HOURS	RIVER	INCREASE
PAPER	MORAL	BED	BODY		

Low frequency words

ZOMBIE	YOGA	YODEL	WEED	VIGIL	VEAL
UNHURT	TONIC	TIDY	TAMPER	SVELTE	SURF
SULTRY	RUBBLE	RODEO	REFUTE	PUN	POSEY
POUT	ORB	NUDISM	ORPHAN	NOSTRIL	NIL
MURAL	MOULD	MEMO	LUGER	LOIN	LOB
KITE	KIMONO	KNEAD	JOUST	JAB	INANE
IMPUTE	HUMID	HYAENA	HOPS		

High frequency control words

BELIEVE	CHILD	CHURCH	COMPANY	DOOR	FACE
FACT	FATHER	MONEY	MORNING	OFFICE	PARTY
PEACE	POWER	PRESSURE	PRIVATE	PROCESS	QUESTION
RETURN	SCHOOL	SPECIAL	TEETH	TIME	WEEK
CLOSE	FORCE	FREE	FRIEND	FUTURE	HEART
LAND	LEVEL	LIGHT	MIND	PAST	STAGE
SYSTEM	TABLE	VALVE	WIFE		

Low frequency control words

BRUNT	CASCADE	CHASSIS	CHIVE	CLOUT	CRANK
DETER	ELAPSE	GLUM	LATHE	LOSER	MALT
NOUN	PEGGED	PERCH	PUTTY	PUZZLER	RAFTER
SLIMMER	SNOOP	THONG	TINT	TOXIN	WHIFF
COVET	CRUTCH	FAUNA	FLAIL	LAZE	LIMPID
LIVID	LOBULE	MEMENTO	MENIAL	OMIT	SARI
SPLICE	USURP	VERTEX	LITHE		

Animal name targets

PIG
LIZARD
BEETLE
SEALION
OCTOPUS
SPIDER
BOAR
GORILLA
TIGER
RAT
FOX
DEER
SQUIRREL
SNAKE
JAGUAR
PIGEON
DOG
FROG
SLOTH
OSTRICH
LION
COW
CAMEL
TOAD
LEOPARD
TORTOISE
HORSE
KOALA
SHREW
STAG
BEAVER
BEE
SHARK
WASP
EMU
BADGER
MOUSE
WORM
TAPIR
MOOSE
CHIMP
PANDA
SEAL
HEDGEHOG
BUDGIE
CAT
MONKEY
HYAENA
HAMSTER
ROBIN
KANGAROO
SPARROW
STOAT
RABBIT
MINK
RHINO
OTTER
ZEBRA
PARROT
SQUID
CRAB
FLEA
LOBSTER
WALLABY
YAK
GIBBON
HIPPO
WHALE
ELEPHANT
ORANGUTAN
PORCUPINE
ARMADILLO
ALLIGATOR
COCKROACH
DOLPHIN
CROCODILE
BUSHBABY
WEASEL
CHICKEN
KESTREL