

---

Centre for Cognitive Science  
University of Edinburgh  
2 Buccleuch Place  
Edinburgh EH8 9LW  
United Kingdom



## **Distinctiveness of Spoken Word Context Predicts Visual Lexical Decision Time**

Richard Shillcock  
rcs@cogsci.ed.ac.uk

Scott McDonald  
scottm@cogsci.ed.ac.uk

Peter Hipwell  
petehip@cogsci.ed.ac.uk

Will Lowe  
will@cogsci.ed.ac.uk

Research Paper EUCCS-RP-1998-4

### **Abstract**

We review various dimensions along which words differ and which, sometimes as part of a word recognition model, have been claimed to predict performance in the visual lexical decision task. Models of word recognition have typically involved inadequate, or non-existent, semantic representations and have dealt with words existing in isolation from any context. We propose an alternative perspective in which it is the relationships between words - reflecting usage and meaning - rather than the discrete entities themselves, that are fundamental to lexical processing. We present Contextual Distinctiveness (CD), a corpus-derived measure of the plurality of the different content-word contexts in which a word occurs in speech, and demonstrate that it is a significant predictor of response times in a simple visual lexical decision task. We argue that LDT effects previously attributed to Age of Acquisition and word frequency should be reinterpreted in terms of CD. As well as subsuming a number of other lexical variables, we detail further advantages of CD, in terms of computational tractability, objectivity, relation to real language, and relation to formal linguistics.

---

# 1 Introduction

The lexical decision task (LDT) is one of the foundational tasks in experimental psycholinguistics: subjects see isolated words or nonwords appear on a screen and are required to make a response - word or nonword, yes or no - as quickly as possible. The LDT has been incorporated into other tasks, such as the cross-modal priming task or the probe latency task, and closely related tasks such as word naming or auditory lexical decision have often been assumed to share processes with the LDT. The recognition of a single word, visual or auditory, would seem to be a basic operation of the human language faculty, and numerous models of word recognition have been developed to characterise the process; these models range from partial, ordinary-language descriptions of the relative importance of the different dimensions along which words vary - the role of concreteness or word frequency, for instance - to implemented computational models that give quantitative data for the ease of recognition of the words forming a substantial part of the lexicon of English (e.g., Seidenberg & McClelland, 1989). In the current paper, we review the role claimed for various factors in the recognition of visually presented single words, and we describe a new measure, Contextual Distinctiveness (CD), that accounts for a significant proportion of the variance in the LDT, and which we claim subsumes a number of these other dimensions.

## 1.1 Dimensions of lexical variation

The literature contains a number of measures of lexical variation that have been demonstrated to be related to ease of processing in particular tasks (principally single-word tasks) and/or to profiles of acquisition or impairment: these measures include Word Frequency (Howes & Solomon, 1951; Whaley, 1978; Monsell, 1991), Age of Acquisition (Gilhooly, 1984), Ease of Predication (Jones, 1985), Concreteness-Abstractness, Imageability (Paivio, Yuille & Madigan, 1968; Paivio, 1971; James, 1975), Familiarity (Gernsbacher, 1984), Standaloneness (Taft, 1994), Polysemy (Jastrzembski, 1981). Some of these measures, such as Ease of Predication (EoP), are based on studies of the intuitions of native speakers; others, such as Word Frequency (WF), are the result of objective measurements made on language corpora. Although each measure embodies a theory about the relevance of that particular dimension to processing and storage, many of the measures are confounded one with another; for instance, WF is typically highly correlated with Age of Acquisition (AoA) (e.g.,  $r = -0.52$  in Whaley's (1978) study). It is therefore necessary to demonstrate both that a particular variable contributes uniquely to determining the variance of the behaviour in question, and that it has explanatory power. Thus, for instance, Baddeley, Ellis, Miles and Lewis (1982, p. 196) state that Imageability has virtually no explanatory power, despite its apparent contribution to various analyses of behaviour, because it is not clear how to model its role in reading<sup>1</sup>. We propose, below, a number of principles by which these intercorrelated variables may be evaluated.

The first criterion we suggest is behavioural prediction: being a powerful predictor of behavioural data is critical, as in various demonstrations that AoA gives a better account of the variance in word naming data than does WF (see Morrison & Ellis (1995) for a review of studies comparing the two variables). The second criterion we suggest is computational tractability: a measure is attractive if it can be modelled computationally, as this allows the measure to be explored more comprehensively and, ultimately, might suggest some neuropsychological instantiation. WF has been an attractive vari-

---

<sup>1</sup>In fact, since the remarks on Imageability by Baddeley et al. were made, the effects of Concreteness have been modelled by Plaut and Shallice (1994), using a connectionist model in which concrete words were defined as possessing more elaborate semantic-feature descriptions than abstract words, demonstrating that initially intuitive measures may be made computationally tractable given certain additional assumptions.

able precisely because it falls naturally out of the Hebbian learning embodied in many current connectionist models of word recognition, although WF has been incorporated into many pre-connectionist models of word recognition in an ad hoc manner. If a measure is computationally tractable then we might expect to find comparable measures at other levels of linguistic description; thus, frequency has been suggested as a predictor of phonological acquisition (Ingram, 1988). This last observation militates against variables such as Imageability which has no interpretation at sublexical levels. The third criterion we suggest is one of objectivity: WF is derived from real language corpora, and is an objective measure, unlike EoP or Concreteness, for example, which both rely on speakers' intuitions. AoA, typically also derived from adults speakers' intuitions, derives further legitimacy from studies such as that by Gilhooly and Gilhooly (1980) showing a close relationship between rated AoA and the objective observation of vocabulary acquisition. The fourth criterion is one of relationship to real language: we can legitimately expect that semantic processing will reflect real world usage, and therefore any relationship that a semantically oriented measure has with real spoken language (the prior and dominant mode of language processing) over and above written language may be counted as a positive feature of a measure. The fifth, and final, criterion we suggest concerns relationship to formal linguistics: if a particular distinction – between words of one type and words of another – is important for processing, we might also expect to find that distinction appearing in formal linguistic analyses, assuming transparency between grammar and processor. Thus, EoP derives some legitimacy from the fact that it reflects established effects of syntactic category, in particular the distinction between content and function words. We return to these criteria below, with particular reference to the novel measure we explore in the current paper.

## **1.2 Problems of, and solutions to, modelling semantic processing**

Some of the dimensions listed above - EoP and Concreteness, for instance - are concerned to some extent with the semantic representation of words. The goal of language is the transfer of meaning, and we might expect meaning to feature prominently in explanations of language behaviour. However, modelling semantic processing presents deep problems for both contemporary Psychology and Philosophy. The developmental model of word recognition proposed by Seidenberg and McClelland (1989) illustrates one solution to the problem: the corner of their model labelled “semantic representations” is left uninstantiated, but is still able to play a role in qualitative explanations that involve a semantic route between the model's orthographic and phonological representations. A second solution to the problem is found in existing feature-based models. For instance, Plaut and Shallice (1994) adopt 68 semantic feature representations (e.g., +/- edible) to model the semantic processing of a small number of words, yet are able to show that their connectionist model robustly produces the constellation of apparently disparate behaviours found in deep dyslexia. McRae, de Sa and Seidenberg (1993) use representative intuitions from speakers to inform their feature-based connectionist model and to demonstrate the effect of correlations between properties in definitions. Gaskell and Marslen-Wilson (1995) employ random vectors to act as the semantic representations in their connectionist model of word recognition. The authors of these last three studies are explicit that the semantic representations they propose are not intended to be adequate ones, merely to reflect qualitatively some of the ways in which speakers/listeners seem to process language.

Psycholinguistic theory has advanced considerably by adopting the convention that lexical representations are discrete entities, and that the semantics of a word can be represented by a simple local representation or by a particular listing of semantic features. In reality it is not possible to provide discrete, necessary and sufficient representations of the meanings of words; easily interpreted, but theoretically problematic expressions like “fake gun”, “stone lion” or “former senator” show that any of

the potential defining characteristics of an entity may be inapplicable, and for a category like “game” it seems that it is more appropriate to talk in terms of particular instances of games having a “familial resemblance” rather than conforming to some discrete definition of “game”. It is possible to conclude that the meanings of words are determined by their contexts of usage (see, e.g., Wittgenstein, 1958), and that a theory of meaning should be holistic, emphasising the interconnectedness of semantic attributions (e.g., Davidson, 1986; Quine, 1960); in a holistic theory of meaning the claim is that the meaning of each word is dependent on the meanings of many other words in the language (Fodor & LePore, 1992).

The problematic nature of putative discrete semantic representations of words is relevant to some of the dimensions of lexical variation discussed above. If discrete semantic representations were possible, then for a simple localist representation of chair containing just such a semantic representation one might log every occurrence of chair and lower some recognition threshold for the word accordingly, and hence finish up with a model of how WF might affect access to word meaning. The fact that such discrete semantic representations are not feasible has meant that the effects of WF have been modelled in terms of non-semantic (orthographic or phonological) representations. In non-connectionist, traditional models of word recognition, there has been a concentration on “identification” in the lexical decision task, prior to the access of meaning. In connectionist models, response strategies have been suggested whereby, for instance, subjects may base their responses to a word in a LDT on the ease with which the orthographic representation of that word can be recreated (Seidenberg & McClelland, 1989). An adequate representation of some aspect of meaning might reform such models at the expense of response strategies or of notions such as identification. The intractability of adequate semantic representations has left other semantically oriented dimensions, such as Concreteness, Imageability, EoP or Polysemy, largely reliant on speakers’ intuitions.

However, it is possible for a model of word recognition to trade on the observation that context is critical to word meaning to model certain aspects of lexical processing. In doing this modelling, the claim is that although the context-based measures to be used certainly do not constitute complete, psychologically realistic semantic representations of words, they do at least approach the essential nature of word meaning from the right direction. Contextual and holistic theories of word meaning are intuitively appealing, but have long been plagued by underspecification and technological constraint. However, the context of usage has been operationalised, in a number of corpus-based studies that seek to explore semantic processing, in terms of a vector based on the lexical context of all the tokens of each word type in a language corpus. This approach is more attractive than the solutions to the problem of modelling lexical semantics considered above, in which the vectors were randomly determined or were the result of intuiting abstract semantic categories like +/- edible. The vectors objectively derived from a corpus are used to construct a high-dimensional hyperspace, in which the usage of the word “cat”, for instance, is represented by a point in that hyperspace. Lund, Burgess and Atchley (1995) employ such a hyperspace representation to explore the difference between semantic and associative priming, relating response-time data to distances in the corpus-derived hyperspace. Hyperspace representations of word meaning are attractive in terms of the criteria we advanced above, in that they have a close relationship to computational architectures and they may be derived objectively from real language corpora. The context-based modelling that we describe below will incorporate and develop existing practice in the construction of such hyperspaces to explore lexical representation.

Hyperspaces are attractive modelling metaphors. In Plaut and Shallice’s connectionist model of deep dyslexia (using simple semantic features on a small scale), recurrent activity allows trajectories to be traced in the hyperspace, in which an approximate representation of the semantics of cat falls into a “point attractor” representing the exact semantics defined for that word. Some of the characteristics of such hyperspaces are counterintuitive, in particular the fact that seemingly disparate entities – such

as cat and cot – may actually be adjacent on some of their dimensions.

We suggested above that a context-based, computationally tractable measure of lexical processing was desirable; we also observed that any particular measure might be further legitimised by a demonstration of its utility at a qualitatively different level of representation. In fact, such a demonstration of the importance of a measure of the diversity of immediate context exists at the phonological level in a recent study by Shillcock and Westermann (in press) concerning phonological acquisition. This study served as the starting point for the research reported in the current paper, and emphasises the role of context in the definition of linguistic representations and in the modelling of language processing. Shillcock and Westermann show that the reported order of acquisition of the consonants of English (Grunwell, 1985) is predicted by phonotactic range, a measure of the diversity of the immediate phonological contexts in which a particular segment may occur: the larger the phonotactic range, the earlier the acquisition. Thus, /t/ has the largest phonotactic range of all the consonants, and it is among the earliest segments to be acquired; it occurs in contexts such as /t'/, /ti/, /tr/, /It/, /lt/, /st/. Phonotactic range is a better predictor than simple segment frequency, suggested by Ingram (1988) on the basis of a very limited crosslinguistic diary study as an important predictor, and frequency makes no contribution independent of phonotactic range. Frequency and phonotactic range are naturally highly correlated and Shillcock and Westermann argue on theoretical grounds for the priority of phonotactic range by demonstrating that it is conceptually equivalent to the learning of phonological constraints. The measure we describe below was an attempt to apply to the lexical level a similar measure of the variability of the immediate context. We began by investigating the lexical type/token ratio in a narrow window either side of the critical word (c.f., Fung, 1995<sup>2</sup>) - a series of explorations that we do not elaborate here - before developing the measure that we describe below, which gave the best modelling performance overall.

### 1.3 The measurement of Contextual Distinctiveness (CD) at the lexical level

Compared with most previous modelling of lexical processing, the emphasis here moves from discrete, isolated entities and their essential properties to the relationships between entities, and their mutual interdependence. The emphasis also moves to a more intuitively appropriate concern with language as it naturally occurs: visual and spoken word recognition typically occurs in contexts of various kinds (even occasionally where that context is the “null context”), and typically involves the extraction of meaning. Although it has been argued that some aspects of visual word recognition performance can be modelled in terms of transcoding purely within a visual domain, as in the response strategy discussed above, this interpretation may reflect more the absence of realistic semantic representations (in any model of word recognition). One of our aims is to contribute towards the rehabilitation of word meaning in a quantitative account of reaction time data. Similarly, although some existing models may incorporate the notion of context – some versions of Becker’s (1980) activation-verification model allow context to suggest candidates, for instance – the mechanisms are underspecified.

Following an investigation of parameters such as window-size and vector size, we define Contextual Distinctiveness (CD) over a window of two words before and two words after the critical word (see also, Huckle, 1996), this calculation being performed on 10.3 million words from the spoken language part of the British National Corpus (BNC). Within this window we exclude from consideration the function words, such as of, is, he than, was, in, and so on; these words numbered 154

---

<sup>2</sup>Fung used a measure of “context heterogeneity” to find, for an English target word, a candidate list of possible Chinese translations. The “context heterogeneity” of an English target word was calculated from a corpus of English text; it was proposed that candidate Chinese translations should have similar heterogeneity values derived in an analogous manner. Fung’s results indicate that the contextual characteristics of word meanings are comparable across languages.

types. We also exclude from consideration very low frequency content words, defined as any content words other than the 2000 most frequent content words. These restrictions correspond to the intuition that neither the very high frequency words or the very low frequency words are able to provide much information that can characterise the semantic context of a word. Finally, the corpus was lemmatised by replacing each inflected token with its canonical form from the CELEX lexical database<sup>3</sup>: for instance, talks, talking, talked, were counted in with their base-form, talk. CD was thus calculated from lemma co-occurrence frequencies, rather than conventional wordform co-occurrence counts. This practice reflects Bradley's (1978) demonstration that an aggregate word-frequency count across transparently inflected forms was a better predictor of response times than was simple frequency of the uninflected form, interpreted as implying that subjects were contacting the representation of a lemma. Lemmatisation is also consistent with the exclusion of the function words; bound functor morphemes are treated identically to freestanding functor morphemes.

The contextual representation of a word can be conceptualised in terms of a multidimensional space (see, e.g., Huckle, 1996; Lund, Burgess and Atchley, 1995). In this type of model, a word corresponds to a vector, or point in a hyperspace, where each dimension of the hyperspace corresponds to some aspect of context. In the current model, each component  $i$  of target word vector  $t$  encodes its co-occurrence count of context word  $c_i$  observed at any position in the limited window. A word vector is thus a superimposition of the range of contexts in which that word occurs.

Next, the vector representing the chance co-occurrence of each context word  $c_i$  with the target word  $t$  was constructed from their independent probabilities of occurrence in the corpus. These probabilities are estimated from corpus frequency counts; the probability of target word  $t$  and context word  $c_i$  co-occurring independently is therefore the product of  $P(t)$  and  $P(c_i)$ . The estimated chance co-occurrence of the pair is simply the independent co-occurrence probability multiplied by the total number of tokens in the corpus<sup>4</sup>.

A straightforward measure for calculating the distance between two vectors in  $n$ -dimensional space is Euclidean distance. We define Contextual Distinctiveness as the distance between a word's actual and chance representations in hyperspace; the larger the distance, the more diverse and distinctive the contexts in which the word occurs, and hence the greater the CD value. Finally, the measure is linearised by taking the natural log.

This procedure produced for each word a CD score ("LogCD" in our analysis below) between 4.17 and 8.50 for the content words considered below; these scores are listed in Appendix A for each of the words used in the experiment below. Intuitively, CD might be expected to trade on many of the dimensions of lexical variation mentioned above. CD is closely correlated with frequency - the more tokens there are of a word, the more opportunity there is for diverse contexts to occur - and hence to AoA. CD might also be expected to capture, objectively, something of what Jones' (1985) EoP was designed to measure, albeit intuitively - the ease with which individual words summon semantic predicates; we might conclude that a word with relatively diverse lexical contexts might be involved in correspondingly many semantic predicates. Diverse lexical contexts might also be expected to signify many shades of meaning for a word, thus a very low CD score (closer to chance) might tend to correspond to a relatively polysemous word, which might have a role as a noun and as a verb (e.g. move), whereas a very high CD word might be one that has, for instance, a limited range of occurrences in idiomatic expressions (e.g., sake). Finally, CD might be expected to be correlated with Concreteness: a concrete word (e.g., hand) may be used metaphorically whereas an abstract one (e.g., duty) cannot, which means that words with a concrete meaning will appear more widely, sometimes

---

<sup>3</sup>CELEX Lexical Database of English (Version 2.5). Dutch Centre for Lexical Information, Nijmegen.

<sup>4</sup>In fact, the term  $wN$ , as opposed simply to  $N$ , allows for the role of window-size, which gives closely similar results.

in a concrete sense, sometimes in an abstract sense, lowering the CD score.

We tested two principal hypotheses. The first hypothesis was that CD would correlate highly with EoP, providing an objective elaboration of that measure. EoP scores were derived from Jones (1985). The second hypothesis was that CD would prove to be a good predictor of response-time data from the simple visual lexical decision task described below, and in particular that it would be a better predictor than EoP and the WF and AoA variables discussed above. AoA for the stimulus words studied in the visual lexical decision task was determined as described below.

## **2 Age of Acquisition study**

Estimates of Age of Acquisition were obtained for the 84 stimulus materials used in the lexical decision task described below. The AoA ratings were obtained using a 7-point scale ranging from 1 (learned at the age of 2 or younger) to 7 (learned at the age of 13 or older) in 2-year increments, as described in Morrison and Ellis (1995). Twelve subjects, drawn from the postgraduate and postdoctoral population of the University of Queensland, were asked to record the age at which they believed they first learned each word either in a spoken or written form. Two data points were lost due to subject error. The means are given in Appendix A.<sup>5</sup>

## **3 Experiment**

A visual lexical decision experiment was conducted to provide response-time data to be modelled by the measures discussed above. CD is calculated from contextual information, meaning that a task involving the recognition of isolated words, with only a “null” context, should be a conservative test of the CD measure. The words used ranged in length and in syntactic category, militating against subjects acquiring any particular sets or strategies in responding.

### **Method**

#### **Stimuli**

Eighty-four words were used in the experiment, plus the same number of orthographically legal non-word items derived from real words. The words used came from a range of syntactic categories: nouns, adjectives, verbs and function words (see Appendix A). The words were a large subset of words taken from Experiment 2 (Jones, 1985), in which the goal of the study was to explore EoP as it ranged across different semantic/syntactic types; the criterion for selecting this subset, and relating to an unpublished study, was that each word also appeared in a text corpus taken from the Wall Street Journal. The non-words were constructed by taking, for each item in the word list, the next word of equivalent length in Francis and Kucera (1982), and altering a single letter, while retaining plausible English orthographic patterning.

#### **Procedure**

The subjects were presented with a fixation point in the centre of a Macintosh LC-475 screen for 500 msec, followed by a stimulus item. The order of presentation for stimulus items was randomised for

---

<sup>5</sup>In scoring the ratings, a judgement of “2 yrs or below” was scored as 2, a judgement of “5-6 yrs” was scored as 6, and so on up to “13 years or more”, which was scored as 13.

Variable	RT	LogLF	WL	AoA	LogCD
Min	467.20	4.37	3.00	2.00	2.58
Max	634.00	11.03	10.00	11.33	8.07
Mean	543.80	7.00	5.60	6.34	4.83
s.d.	37.26	1.24	1.77	2.27	1.20

Table 1: Descriptive statistics for the variables (LogLF is derived from the spoken language corpus).

each subject. All subjects saw all of the stimulus materials. The task was to decide whether each item was a word or non-word: recognition was signalled with the index finger of the dominant hand, non-recognition with the other hand. Response times were taken with the button box produced for use with the Psyscope program, which was used to run the experiment. As soon as the subject responded, the word was replaced with a new fixation point. A short practice run, consisting of 10 items not used in the analysis was given to familiarise subjects with the task.

## Subjects

There were 12 subjects, ranging in age between 20-30. They were all volunteers: no payment was given for participation. All had English as their first language, no known language deficits, and normal or corrected-to-normal vision. Two subjects were left handed.

## Data Analysis

If subjects failed to respond within 2000 msec, then the next item would be displayed: this never occurred. The error rate for responses to words was 2% overall (ranging from 0-6 errors per subject): where an error occurred, we substituted the average response time for that subject. From the resulting figures, we calculated the overall mean response times for each word. These response times were analysed as the variable RT in the regression analyses described below.

## 4 Results of the regression analyses

In the following analyses we are concerned only with the data from the 65 content words; we return, below, to the distinction between content and function words. We will also explore the distributional statistics provided respectively by the spoken and written language data of the BNC.

First, we test our first hypothesis, namely that there would be a significant relationship between EoP and CD (or rather, LogCD, the log-transformed CD) with the implication that both would be good predictors of RT. For all 84 items, there was a significant negative correlation between EoP and LogCD ( $r = -0.603$ ,  $p < 0.0001$ ), but for the 65 content words alone, there was no significant linear relationship ( $r = -0.062$ ). We conclude that the significant correlation for all the materials was due solely to the function-/content-word distinction. Further, there was no linear relationship between EoP and mean RT:  $r = -0.008$  for the 65 items;  $r = 0.118$  for all 84 items. We conclude that EoP is not a significant predictor of speeded LDT data, and that CD is not usefully construed as an objective elaboration of the EoP measure. We devote the rest of the analysis to the behaviour of LogCD and its relationship with the other variables.

Table 1 gives descriptive statistics for the remaining variables explored. Table 2 shows the correlations between AoA, LogCD, Word Length (WL), LDT response time (RT) and LogLF (the log-transformed lemma frequency in the spoken corpus). (Our use of lemma frequency in preference to

Variables	1	2	3	4	5
1.RT	1.000				
2.WL	.293**	1.000			
3.LogLF	-.293**	-.235*	1.000		
4.AoA	.307**	.504***	-.356**	1.000	
5.LogCD	-.393***	-.292*	.915*	-.439***	1.000

Table 2: Correlation matrix of the four predictor variables with response times, for the spoken language corpus. Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .005$ , (one-tailed).

Predictor	beta	<i>t</i>	<i>df</i>	<i>p</i>
LogLF	0.356	1.248	61	.217
WL	0.152	1.142	61	.258
AoA	0.076	0.532	61	.600
LogCD	-0.641	-2.156	61	.035

Multiple  $R^2 = .2153$ ,  $F(4,60) = 4.12$ ,  $p = .0052$

Table 3: Summary of simultaneous regression analyses for the spoken language corpus.

frequency of the individual wordform was discussed above; this choice was confirmed in analyses of the current data, which we do not present here.)

As Table 2 shows, several of the variables are intercorrelated; RT was significantly correlated with WL, LogLF, AoA and LogCD ( $p = .009$ ,  $p = .009$ ,  $p = .006$ ,  $p = .001$ , one-tailed, respectively). Nevertheless, LogCD possessed the highest correlation with RT, despite the fact that, as Table 1 shows, LogCD had the smallest range of any of the variables. Multiple regression analysis allows us to partial out the influence of potentially confounding variables. We conducted a simultaneous multiple regression analysis on RT, with LogLF, WL, AoA and LogCD as predictors. Table 3 summarises the results.

The results of the analysis indicated that once the variance from the other variables had been taken into account, LogCD emerged as the best predictor of RT ( $p = 0.035$ ). The effects of WL, AoA and LogLF were clearly nonsignificant when LogCD was in the equation. However, the high intercorrelation between LogLF and LogCD ( $r = .915$ ) indicated that the simultaneous regression model is not appropriate for determining which of the two is the better, independent predictor of the variance of RT. We therefore conducted further analyses to distinguish the relative contribution of each of the predictor variables. First, we determined the effect of partialling out combinations of the predictor variables. When WL was held constant, the correlation between LogCD and RT remained significant ( $r = -.336$ ,  $p = .003$ ). The correlation between LogLF and RT was also significant ( $r = -.241$ ,  $p = .028$ ), and the correlation between RT and AoA was marginally significant ( $r = .193$ ,  $p = .064$ ) with WL partialled out. Holding both WL and AoA constant, LogCD was still significantly correlated with RT ( $r = -.292$ ,  $p = .01$ ), but there was only a marginally significant correlation between LogLF and RT ( $r = .198$ ,  $p = .06$ ). Controlling additionally for LogLF still resulted in a significant correlation of LogCD with RT ( $r = -.268$ ,  $p = .018$ ). When variance due to WL, AoA and LogCD was partialled out, however, the correlation between LogLF and RT did not reach

significance ( $r = .159, p = .2176^6$ ).

Next, a hierarchical multiple regression analysis was performed on RT, with WL, LogCD, AoA and LogLF as predictors. A similar result was found as with the partial correlation analyses. When only WL was included as a predictor, an  $R^2$  of .086 was obtained. Adding LogCD increased  $R^2$  to .189, a significant increase,  $F(1, 62) = 7.89, p < .01$ . The further addition of either LogLF or AoA resulted in a nonsignificant increase in  $R^2$  (to .212,  $F(1, 61) = 1.73$ , n.s., and to .195,  $F(1, 61) < 1$ , respectively). A regression analysis starting with WL and LogLF only gave an  $R^2$  of .139. The increase in  $R^2$  from the regression result with WL as the only predictor was marginally significant,  $F(1, 62) = 3.82, .1 > p > .05$ . Adding AoA raised  $R^2$  to .155, which was not a significant change,  $F(1, 61) = 1.11$ , n.s.. The difference between the RT variance accounted for by WL, WF and AoA and the variance accounted for by the addition of LogCD to the equation was significant ( $F(1, 60) = 4.65, p < .05$ ) resulting in an  $R^2$  of .215.

From the results of both the partial correlation and the hierarchical regression analyses, it appears that LogCD is more predictive of lexical decision latencies than either LogLF or AoA. LogCD makes a larger, independent contribution towards accounting for RT variance than does LogLF or AoA. The optimal combination of the four predictor variables considered here for predicting visual lexical decision latencies includes word length and Contextual Distinctiveness, and excludes word frequency and Age of Acquisition.

The sign reversal of the standardised regression coefficient (Beta) for LogLF indicated that suppression is taking place between LogLF and LogCD; it appears that one function of LogLF in the multiple regression equation is to suppress a portion of the variance in LogCD that is irrelevant to response times<sup>7</sup>. This result is consistent with our view on the secondary role of word frequency: LogCD subsumes nearly all of the role of LogLF, while uniquely accounting for 6.08% of the RT variance. In comparison, the unique contributions from WL, LogLF and AoA are 2.04%, 1.71% and 0.37% respectively.

*Calculating CD from a written corpus.* In order to compare the performance of LogCD across spoken and written genres, we created a sub-corpus of the written language part of the BNC comparable in size to the spoken sub-corpus<sup>8</sup>. LogCD was calculated from co-occurrence counts from the written sub-corpus in exactly the same manner as described above.

Correlation coefficients were computed between RT, WL, LogLF, AoA and LogCD (see Table 4). RT was significantly correlated with WL, LogLF, AoA and LogCD ( $p = .009, p = .026, p = .006$ , and  $p = .010$ , one-tailed, respectively).

A simultaneous multiple regression analysis was carried out on RT, with LogLF, WL, AoA and LogCD as predictor variables (see Table 5). None of the predictors accounted for a unique, reliable proportion of the RT variance. This was due to confounds between pairs of the variables: LogCD and LogLF are highly intercorrelated ( $r = .864$ ) as are AoA and WL ( $r = .504$ ). In order to determine the relative importance of the individual variables, further analyses were conducted.

We carried out partial correlation analyses using combinations of the four predictor variables.

---

<sup>6</sup>In this instance we report a two-tailed probability, given that the direction of the relationship unexpectedly switched.

<sup>7</sup>With regard to the issue of whether LogCD is suppressing LogLF, thereby causing the role of LogLF to be underestimated, note that the positive relation between RT and word frequency, indicated by the Beta of +0.356 for the relation between RT and LogLF, is inconsistent with the usual effects of frequency. It does not support a word frequency effect to say that LogCD was suppressing a positive relationship. Note also that the partial correlation between LogLF and RT, with all other variables held constant, is also positive (+0.159).

<sup>8</sup>Since the written part of the BNC is approximately 90 million words distributed among 3209 texts, in order to obtain a representative sample, we constructed a 10.3 million word sub-corpus from a random selection of individual texts.

Variables	1	2	3	4	5
1.RT	1.000				
2.WL	.293**	1.000			
3.LogLF	-.242*	-.104*	1.000		
4.AoA	.307***	.504***	-.133**	1.000	
5.LogCD	-.288**	-.109	.864***	-.118	1.000

Table 4: Correlation matrix of the four predictor variables with response times, for the written language corpus. Note: \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .005$ , (one-tailed).

Predictor	beta	$t$	$df$	$p$
LogLF	0.059	0.253	61	.802
WL	0.169	1.247	61	.217
AoA	0.195	1.431	61	.158
LogCD	-0.297	-1.281	61	.205

Multiple  $R^2 = .1807$ ,  $F(4,60) = 3.31$ ,  $p = .016$

Table 5: Summary of simultaneous regression analyses for the written language corpus.

With WL partialled out, LogCD and RT were significantly correlated ( $r = -.269$ ,  $p = .016$ ). When WL was held constant, the correlation of LogLF and RT was significant ( $r = -.222$ ,  $p = .039$ ), but that between AoA and RT was only marginally significant ( $r = .193$ ,  $p = .064$ ). With WL and AoA partialled, LogCD remained significantly correlated with RT ( $r = 0.261$ ,  $p = .020$ ), as did LogLF ( $r = -.209$ ,  $p = .050$ ). When WL, AoA, and LogLF were partialled out, the correlation between LogCD and RT was not significant ( $r = -.163$ ,  $p = .103$ ). Partialling out WL, AoA and LogCD resulted in a nonsignificant relationship between RT and LogLF ( $r = .033$ , n.s.).

We also conducted a hierarchical multiple regression analysis on RT, with WL, LogCD, AoA and LogLF as predictor variables. The results were similar to those obtained in the partial correlation analyses. WL as the sole predictor gave an  $R^2$  of .086. The addition of LogCD to the regression equation increased  $R^2$  to .152, a significant increase,  $F(1, 62) = 4.85$ ,  $p < .05$ . The further addition of either LogLF or AoA to the equation resulted in a nonsignificant increase of  $R^2$  (to .153,  $F(1, 61) < 1$ , and to .180,  $F(1, 61) = 2.04$ , n.s., respectively). The regression analysis with WL and LogLF as predictors resulted in an  $R^2$  of .131, which was a marginally significant increase ( $F(1, 62) = 3.21$ ,  $.1 < p < .10$ ) from the  $R^2$  when WL was the only predictor. Adding AoA gave an  $R^2$  of .158, a nonsignificant increase,  $F(1, 61) = 1.97$ , n.s.. The further addition of LogCD to the equation did not significantly raise  $R^2$ :  $F(1, 60) = 1.64$ , n.s.

The results of the partial correlation and the hierarchical regression analyses suggest that word frequency and Contextual Distinctiveness calculated using texts from the written portion of the BNC are approximately equally predictive of lexical decision latencies, although the above analyses were unable to disentangle the confounded variables, LogCD(written) and LogLF(written). The difference between the regression analyses based on the different genres, indicating that CD is a better predictor of lexical decision times than word frequency only for a spoken language corpus, suggest the primacy of spoken language data in language processing. We return to this point below.

## 5 Discussion

The results show that the Contextual Distinctiveness measure is a powerful predictor of performance in the simple visual lexical decision task. The results also show that, in terms of predictive power, CD compares favourably with the other variables we have discussed. “Behavioural prediction” was the most important of the five criteria that we suggested above as means of distinguishing between the various lexical dimensions that have been put forward to model lexical processing. CD is a measure that rates favourably on each of the other four criteria, as we discuss below.

*CD and the criteria for modelling lexical data.* As well as being the best individual predictor of response times in the LDT, of the dimensions considered, and therefore fulfilling the criterion of behavioural prediction, CD is attractive because it is eminently computationally tractable: CD is derived from the high-dimensional hyperspaces that result from a number of statistical and connectionist approaches to modelling behaviour. In addition, CD resembles a predictor of phonological acquisition, discussed above, in that both are measures of locally-defined complexity; such a correspondence between qualitatively different levels of processing tends to add legitimacy to both measures. Third, CD is an objective measure, derived from a real corpus. CD thus directly represents the input that the human language processor receives, at the level of lexical access; in contrast, a measure like Concrete-ness can only be said to match real language input at a very much more abstract level of description. The fourth criterion which we considered above was “relationship to real language”; CD fulfils this criterion by virtue of the superior modelling we report on the basis of the spoken language subset - compared to the written language subset - of the BNC. At first glance, we might predict that text corpora would provide the best model of behaviour elicited by a visual LDT, essentially a text-based task; this prediction implies sensitivity to the modality and/or to relatively peripheral aspects of the stimulus. However, we argue that spoken - as opposed to written - language processing, both in perception and production, is a highly practiced skill which predates learning to read by several years and which is typically more practiced even in highly literate adults. Further, spoken language is implicated in some reading behaviour in the form of subvocal rehearsal and the claimed automatic generation of phonological representations during reading (see, e.g., Rubenstein, Lewis & Rubenstein, 1971); therefore, there may be less opportunity than first appears for the separation of spoken and written processing. Templin (1957) estimates that an English-speaking child will have acquired some 14,000 words by age 6, an age at which reading behaviour is likely to be still fairly rudimentary compared with adult reading behaviour. A substantial functioning semantic space has thus been constructed before any very extensive distributional statistics of text could have been computed. We should expect this semantic space to be the foundation for the adult semantic space and, therefore, for spoken language corpora to reflect processing more accurately than do text corpora. The reliability of text corpora as generators of CD measures is liable to be compromised by the inclusion of a higher proportion of low frequency words, compared with spoken language corpora. Our explorations of CD showed that its predictability peaked at a level at which approximately 2000 words (the 2000 most frequently occurring content words) were used to define the context vectors. We might suggest that such a number of words/dimensions represents some basic goal in the achievement of semantic competence<sup>9</sup>. We found that the 2000 most frequently occurring content words overlapped by 76.95% between the speech and text corpora, and we can only speculate that the unique words/dimensions provided by the text corpus were less semantically discriminating, possibly more idiosyncratic, than the words/dimensions unique to the speech corpus. We might also speculate that the expressive syntactic devices that are more com-

---

<sup>9</sup>However, the precise number would need to be subject to additional cross-corpora checking.

mon in text - the fronting, clefting, unbounded dependencies - serve to confound the narrow-window constraint imposed by our definition of CD: potentially important words/dimensions may have been taken outside of the  $+2/-2$  window in text, but remained within the window in casual speech, which tends to be composed of small clauses with a relative paucity of complex construction types. Finally, the word-by-word progress of processing that is implicit in the narrow window parameter may not be an entirely accurate description of some reading behaviour in which words may be processed partly in parallel, skipped and/or re-read; in contrast, the temporal nature of speech makes lexical processing overwhelmingly word-by-word (Bard, Shillcock & Altmann, 1988), the exceptions largely being the function words that are irrelevant to CD. We conclude that CD's sensitivity to speech/text differences is a mark of legitimacy that should be extended as a criterion of other dimensions of lexical variation.

To an extent, the criterion of "relationship with real language" concerns psychological reality. CD has a measure of psychological reality insofar as the small size of the window in the calculation of CD contrasts favourably with the large window used in some past explorations of context in corpora, and confirms other findings of the efficacy of small windows (see, e.g., Lund et al., 1995; Huckle, 1996). The value of a small window onto the linguistic input suggests that CD might be an important dimension in language development, given that infant resources of memory and attention are restricted. In adult language processing, a window of five words is within the scope of the phonological short term store (Baddeley, 1986).

The fifth and final criterion proposed above was that of the relationship with formal linguistics. This relationship is explicit in CD in that we exclude from consideration the function words: these are not employed in defining CD for individual words, nor do we seek to account for LDT performance on function words. At first sight, this may seem a restriction on the universality of CD. However, we follow other researchers who have used corpus statistics to characterise lexical representations (e.g., Lund et al., 1995), but who use numerical rather than linguistic criteria to avoid the swamping effects of the function words by excluding, for instance, the most frequent  $x$  words regardless of type. Because function words dominate the upper extreme of the frequency range, these two approaches are convergent; however, we argue that the linguistically-based criterion is the more psychologically realistic criterion. The difference between content words and function words is a deep distinction in formal linguistics: function words have become more and more important in the projection of syntactic structure in Chomskyan linguistics over the last decade – "noun phrases" have become "determiner phrases", for instance – and there are equally compelling distinctions between the two word classes: in English, the two word classes differ at the level of phonetic realisation (see, e.g., Cutler, 1993), lexical neighbourhood statistics (Shillcock, Kelly, Buntin & Patterson, submitted), priming behaviour (Shillcock & Bard, 1993), vulnerability to impairment (Friederici & Shoelne, 1980) (see Cann, 1996, for a recent review). Function words are able to coerce syntactic-semantic changes in accompanying content words in a way that is not possible in the reverse direction. We should expect that an equally deep division should exist between the two word types in the semantic processing we are currently considering. As with the speech/text distinction discussed above, we claim that the function/content-word distinction embodied in CD is an indication of the legitimacy of that measure, a sign of its psychological reality, and deserves to be exported to other measures. It should be noted, finally, that the function words emerge as a class when CD is calculated, as with a frequency ordered list; they naturally fall out as having very high values of CD. This observation suggests a means by which the separate storage/processing of function and content words might emerge in language acquisition if CD is an important processing dimension at that stage.

In summary, CD is favoured on all of the five criteria we proposed above as means by which a psycholinguistically useful dimension of lexical variation might be defined. The other measures proposed in the literature do not fare so well on these criteria. We go on to discuss the relationship of

CD to the other lexical variables.

*CD and EoP.* We expected CD to be closely related to EoP. Jones intended EoP to represent the ease with which each word might summon semantic predicates. Thus, a high-EoP word like dog can readily be put into statements such as a dog is a type of animal, a dog often lives in a kennel, a dog barks when angry, and so on. Our expectations regarding EoP and CD were based on the prediction that different words occurring close to dog, for instance, should indicate that dog was being involved in a correspondingly large number of different predicates. We found that the significant relationship between EoP and CD depended exclusively on the function words; these words score at the extremes of both measures. Our failure to find either a significant relationship between CD and EoP for the content words, or a role for EoP in modelling the LDT content-word data, suggests that EoP is a relatively abstract measure (borne out by its intuitive nature) with no direct relationship to LDT performance. The absence of a significant correlation with CD (for the content words alone) may perhaps be partly explained by the fact that a high CD score for a word may derive from that word occurring very often in a limited variety of contexts; this constitutes a departure from chance contexts but does not correspond to a large variety of different semantic contexts. We conclude that some of the original intuitions behind EoP may be more successfully expressed in terms of CD, which has the advantage of being a good predictor of online lexical processing. We predict that CD has implications for the modelling of impaired processing, as in deep dyslexia for which EoP was developed. Finally, just as EoP was conceived as a more tractable measure that would account for intuition-based measures such as Imageability, we propose CD as an objective measure that captures semantically oriented measures such as Imageability, and Polysemy.

*CD and AoA.* Morrison and Ellis (1995) have suggested that AoA is a more powerful predictor of language processing than is WF, although they still allow for an independent role for WF in the visual LDT. From the work of Morrison and Ellis, and the previous work that they review, AoA emerges as the strongest candidate of the competing measures of lexical variation; Morrison and Ellis propose the wholesale substitution of AoA for WF in processing explanations of naming behaviour. This claim for the priority of AoA is based principally on its abilities as a predictor of response time in naming and in LDT. In addition, AoA is computationally tractable: Morrison and Ellis echo Brown and Watson's (1987) suggestion that early-learned words configure the initial space for the phonological representation of words and determine the way in which later-acquired words are represented. Finally, AoA possesses a measure of objectivity: intuitions about AoA are closely related to objective observations of acquisition (Gilhooly & Gilhooly, 1980).

We found a significant correlation ( $r = .307, p < .01$ ) between AoA and RT in the LDT; early-acquired words were responded to faster. This result reflects the (somewhat mixed) existing studies of the relationship between AoA and LDT data. A non-significant relationship was reported by Gilhooly and Logie (1982), but a significant independent effect was reported by Whaley (1978), Butler and Hains (1979), and Nagy, Anderson, Schommer, Scott and Stallman (1989). However, in reporting their own results - both AoA and WF exerted significant effects in the LDT - Morrison and Ellis claim that failure to find effects of AoA can sometimes be attributed to the problematic interpretation of stepwise multiple regression results: shared variance between two predictors can lead to the less powerful of the two not being accorded its rightful status when the other variable is taken out of the analysis (see also Morris, 1981). However, sets of stimulus items differ. Although the current stimulus materials were not originally designed to test AoA, they provided a reasonable range of means (from 2.0 to 11.3), with the inevitable bulge in the 3-4 years category. However, the correlation between AoA and LogLF

( $r = .356$ ) is not as high as is sometimes reported in the literature; Rubin (1980), for instance, reports an  $r$  of .40, and Gilhooly and Logie (1982) report one of .71, these differences reflecting the choice of words. We might also note that the current stimulus materials differed from many AoA studies in containing a range of syntactic categories. The content words contained unambiguous adjectives like golden, famous and different, and unambiguous verbs like eat, include and lose. (In addition, our subjects were also required to judge function words; although these responses were not included in our analysis, they still may have affected subjects' behaviour in terms of making some response strategies less attractive). In contrast, the materials used by Morrison and Ellis, for instance, all at least possessed noun readings, and for most the noun reading was the most frequent interpretation. When the nouns in the stimulus materials of the current study were analysed alone, the correlation between AoA and RT fell to  $r = .229$ , from  $r = .307$  for all the content words, and when the function words were included in this correlation  $r$  fell to .182. We may provisionally conclude that the increased range of syntactic categories in the stimulus materials in the current study does not mask a superior role for AoA in modelling responses to the nouns. However, the concentration on nouns in some other studies may have affected subjects' response strategies to the advantage of AoA; our own study, showing a prime role for CD, is psychologically more realistic in that a more representative range of syntactic categories is used.

The discussion above has implied that CD is preferable to AoA on grounds other than predicting behaviour. AoA has some architectural implications: Brown and Watson suggest a "first in, first out" principle in which the phonological space in which the individual's lexicon is represented is conditioned for the rest of the individual's life by early exposure to particular words. Morrison and Ellis endorse this conclusion, although they note that it does not sit easily with the gradient descent algorithms standardly used in connectionist cognitive modelling, in which subsequent experience may override early training; they go on to suggest that the self-organising Kohonen network (Kohonen, 1984, 1990) might allow the effects of early experience, demanded by AoA, to be modelled. It does not seem to be possible to convert this claim regarding the importance of early experience to one of the importance of the sum total of lifetime experiences with a particular word; analyses of object-naming speed (Carroll & White, 1973) and word-naming speed (Gilhooly, 1984) suggest that a straightforward AoA measure produces a better prediction than related "lifetime" measures.

This computational interpretation of AoA, developed out of word-naming studies, becomes a problem for the AoA measure in the light of Morrison and Ellis's demonstration that AoA is a significant predictor of visual lexical decision time, as it means that performance in the LDT requires a phonological component: subjects respond to certain words quickly in the visual LDT because the phonological representations of those words were laid down early. Morrison and Ellis cautiously accept this conclusion, and go on to make a prediction that AoA's role might be reduced in a visual lexical decision study that involves pseudohomophones, as this would bias subjects against relying on phonological information in lexical decision.

If we accept that CD should displace both WF and AoA in modelling lexical processing, then we are freed from this dependence on a phonological strategy in responding in the LDT. Certainly, CD is critically derived from spoken - as opposed to text - corpora, but this reflects semantic processing. Although phonological and orthographic information may all have a role to play in subjects' performance in the visual LDT, CD exercises its role directly by means of semantic structure.

Overall, then CD is a better predictor of visual LDT performance than AoA; CD predicts performance robustly on a range of content words belonging to the range of syntactic categories. Its predictions are based on word-meaning and, unlike AoA, do not rely on assumptions concerning phonological representations. CD is more attractive in its computational tractability, and we suggest the wholesale replacement of AoA by CD in discussions of language processing.

*CD and WF.* The results show that LogLF does not survive as a significant predictor of RT in the LDT once LogCD is taken into account. We endorse the claims made by Morrison and Ellis, and others before them, that WF is not a psychologically realistic predictor of lexical processing; the variance in behaviour that it apparently accounts for is best ascribed to another variable. In our case, of course, we wish to say that CD is the best predictor of processing performance. The pervasive control of lexical stimuli in psycholinguistic experiments for word frequency should be replaced by a similar control for CD.

*Ecological validity.* The CD measure is motivated by the belief that processing explanations of psycholinguistic results should be ecologically valid. Specifically, we predict that subject performance in psycholinguistic experiments will reflect the semantic processing that occurs when comprehending ordinary text or speech. Since the natural goal of sentence processing is the extraction of meaning, we assume that linguistic processing in an experimental context involving recognition will be devoted to extracting semantic information. Absolute target word frequency is relevant only to identifying the word and is of limited value in the ongoing process of sentence comprehension; a contextual measure will be more immediately useful, since it constitutes a first step towards the extraction of sentence meaning, and should provide richer data for the development of a successful response strategy in a task such as the LDT.

*Holistic and contextual theories of meaning.* Contextual theories of language use have previously been constrained by lack of computational resources. However, current computing power makes it possible to extract distributional information from large, representative corpora. Holistic theories of meaning are often inspired by Wittgenstein's (1958) dictum: "Don't look for the meaning, look for the use" (Davidson, 1986; Fodor & LePore, 1992 for a review). The dictum implies that word meanings are not entities to which words attach. Rather, a word's meaning is given by the pattern of its uses in connected discourse. Similar dicta can be found in the (now reviving) field of distributional linguistics - "you shall know a word by the company it keeps" (Firth, 1950; see also Harris, 1960, Deese 1965). However, quantitative research has been constrained by the problem of deciding which aspects of language use are important to word meaning. We have approached this problem by assuming that it is only linguistic context that is important. This turns out to be a fruitful idealisation (see also Charniak, 1993). This approach makes testing holistic theories computationally tractable, and also allows us to reinterpret previous holistic theories of meaning in quantitative terms.

*Hyperspace models of meaning.* Holistic models have a natural expression as a hyperspace, with axes defined by the quantities in terms of which individual word meanings are defined. A very holistic theory will define a space of very high dimensionality. Previous hyperspace theories have been criticised for not fully specifying the axes that define the space (Fodor & LePore, 1992). This criticism is particularly problematic if the axes represent quantities that would be used by non-holistic theories, such as perceptual features of objects; in this case the hyperspace model simply reduces to the non-holistic theory. Theories that define word meaning in terms of a co-occurrence vector are immune to this criticism, because the axes of the hyperspace are determined by quantifiable aspects of language structure, not elements of a semantic decomposition.

Consistent with contextual and holistic theories of meaning we assume that word meaning can be represented in a hyperspace, with axes reflecting co-occurrence regularities. We take the predictive success of CD to be an indirect argument for the contextual, holistic theories that inspired it.

*Implications for models of word recognition and the LDT.* As Morrison and Ellis observe, replacing the effects of word frequency by a qualitatively different measure is a severe problem for current connectionist models of word recognition (e.g., Seidenberg & McClelland, 1989), given that frequency effects naturally fall out of the learning exhibited by such models. We propose that such models might still accurately reflect lexical processing if semantic processing is incorporated, by requiring such models to construct semantic hyperspaces equivalent to that explored in the CD measure in the current study. Lexical decision time, as influenced by semantic factors, is determined by the distance from a chance location in that hyperspace, as we have described above: the greater the distance from chance, the more compelling is the “semantic existence” of the input. Further work is required to demonstrate the precise reliance on semantic as opposed to phonological and orthographic factors in the LDT, but our default assumption is that semantic factors are given priority, reflecting the ecological reality of language use in transferring meaning. Nevertheless, in specific sets of stimuli - containing a large proportion of pseudohomophones, or unpronounceable nonwords, for instance - subjects may be sufficiently flexible to base their responses more on orthographic or phonological criteria (see, for instance, Voice, 1995); in these cases the fact that frequency effects fall naturally out of the (more peripheral) conversion of the orthographic input into a phonological representation may still play a part in lexical decision.

The evidence we have presented against the direct role of word frequency, together with previous evidence to this effect discussed above, counts against those non-connectionist models of word recognition that incorporate a frequency effect (e.g., Morton, 1969; Forster, 1976).

## 6 Conclusions

We have presented an account of data from a single experiment, yet the results in favour of our measure of Contextual Dependency, derived from spoken language, are robust and clear. CD emerges as a powerful predictor of lexical processing in this basic task. CD allows us to model subjects’ behaviour in a way that EoP does not. CD eclipses the effects of word frequency, allowing us to conclude that some of the demonstrations of the role of word frequency reported in the literature are due in large part to WF approximating CD. Just as AoA has been shown to subsume WF, so CD apparently subsumes AoA. We suggest that CD is a more cognitively insightful variable than the other dimensions of lexical difference that we have reviewed above, in particular Age of Acquisition (AoA). Many lexical dimensions are closely confounded - CD, WF, AoA, Concreteness, Imageability and others - and it is necessary to have principled criteria to distinguish between them. We have suggested five such criteria: behavioural prediction, computational tractability, objectivity, relationship to real language, and relationship to formal linguistics. We have shown that CD is an attractive measure on all of these counts. We propose that models of lexical processing should be modified to allow a role for CD. This means extending the architectural scope of connectionist models to allow them to construct sufficiently complex hyperspaces trained on connected language, and employing a measure of distance from chance as a criterion for lexical decision. Further, we conclude that spoken language has priority over written language in determining the linguistic representations accessed in tasks such as visual lexical decision. Finally, we conclude in favour of the holistic and contextual theories of meaning that inspired the current work.

## References

- Baddley, A. D., Ellis, N. C., Miles, T. R., and Lewis, V. J. (1982). Developmental and acquired dyslexia: A comparison. *Cognition*, 11:185–199.
- Bard, E. G., Shillcock, R. C., and Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics*, 44:395–408.
- Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Canadian Journal of Psychology*, 8:493–512.
- Bradley, D. (1978). *Computational distinctions of vocabulary type*. Cambridge, MA: MIT.
- Brown, G. D. A. and Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory and Cognition*, 15(3):208–216.
- Butler, B. and Hains, S. (1979). Individual differences in word recognition latency. *Memory and Cognition*, 7:68–76.
- Cann, R. (1996). Categories, labels and types: functional versus lexical. Technical report, Edinburgh University Occasional Papers in Linguistics.
- 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.
- Charniak, E. (1993). *Statistical Language Learning*. MIT Press.
- Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, 22:109–131.
- Davidson, D. (1986). *Inquiries into Truth and Interpretation*. Oxford University Press.
- Firth, J. R. (1968). *Selected papers of J. R. Firth*. Indiana University Press.
- Fodor, J. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Blackwell.
- Forster, K. I. (1976). Accessing the mental lexicon. In Wales, R. J. and Walker, E., editors, *New Approaches to Language Mechanisms*. Amsterdam: North Holland.
- Francis, W. N. and Kucera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin Company: Boston.
- Friederici, A. and Shoelne, P. W. (1980). Computational dissociation of two vocabulary types: Evidence from aphasia. *Neuropsychologia*, 18:11–20.
- Fung, P. (1995). Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus. In *Proceedings of the 3rd Annual Workshop on Very Large Corpora*, pages 173–183, Boston MA.
- Gaskell, G. and Marslen-Wilson, W. D. (1995). Modelling the perception of spoken words. In *Proceedings of the 17th Annual Meeting of the Cognitive Science Society*, pages 19–24. Hillsdale, NJ: Lawrence Erlbaum Associates.

- 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.
- Gilhooly, K. J. (1984). Word age-of-acquisition and residence time in lexical memory as factors in word naming. *Current Psychological Research and Reviews*, 3:24–31.
- Gilhooly, K. J. and Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, 71:105–110.
- Gilhooly, K. J. and Logie, R. H. (1982). Word age-of-acquisition and lexical decision making. *Acta Psychologica*, 50:21–34.
- Grunwell, P. (1985). *Phonological Assessment Of Child Speech (PACS)*. NFER-Nelson, Windsor: UK.
- Harris, Z. S. (1960). *Structural Linguistics*. University of Chicago Press.
- Howes, D. H. and Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, 41:401–410.
- Huckle, C. C. (1996). *Unsupervised categorization of word meanings using statistical and neural network methods*. PhD thesis, Centre for Cognitive Science, University of Edinburgh.
- Ingram, D. (1988). The acquisition of word-initial /v/. *Language and Speech*, 31:77–85.
- James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 1:130–136.
- Jastrzemski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence and the lexicon. *Cognitive Psychology*, 13:278–305.
- Jones, G. V. (1985). Deep dyslexia, imageability and ease of predication. *Brain and Language*, 24:1–19.
- Kohonen, T. (1984). *Self-organization and associative memory*. Springer-Verlag.
- Kohonen, T. (1990). The self-organizing map. *Proceedings of IEEE*, 78:1464–1480.
- Lund, K., Burgess, C., and Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 660–665. Mahwah, NJ: Lawrence Erlbaum Associates.
- McRae, K., de Sa, V., and Seidenberg, M. S. (1993). The role of correlated properties in accessing conceptual memory. Manuscript.
- Monsell, S. (1991). The nature and locus of word frequency effects in reading. In Besner, D. and Humphreys, G. W., editors, *Basic Processes in Reading: Visual Word Recognition*. Lawrence Erlbaum Associates.
- Morris, P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple regression analysis. *Memory and Cognition*, 9:277–282.

- Morrison, C. and Ellis, A. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:116–133.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76:340–354.
- Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., and Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 24:262–282.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart and Winston.
- Paivio, A., Yuille, J. C., and Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, monograph supplement*, 76(1).
- Plaut, D. C. and Shallice, T. (1994). *Connectionist modelling in cognitive neuropsychology: A case study*. Lawrence Erlbaum Associates: Hove, England.
- Quine, W. V. O. (1960). *Word and Object*. MIT Press.
- Rubenstein, H., Lewis, S. S., and Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 10:645–657.
- 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.
- Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96:523–568.
- Shillcock, R. C. and Bard, E. G. (1993). Modularity and the processing of closed class words. In Altmann, G. T. M. and Shillcock, R. C., editors, *Cognitive Models of Speech Processing: The second Sperlonga Meeting*. Erlbaum.
- Shillcock, R. C., Kelly, M. L., Buntin, L., and Patterson, D. (1996). Visual processing of content and function words: similarities and differences in the clarity gating task. Submitted.
- Shillcock, R. C. and Westermann, G. (1996). The role of phonotactic range in the order of acquisition of english consonants. In *Proceedings of the ICPLA, 1996*, Munich.
- Taft, M. (1990). Lexical processing of functionally constrained words. *Journal of Memory and Language*, 29:245–257.
- Voice, J. K. (1995). *Competition processes in visual word recognition*. PhD thesis, Centre for Cognitive Science, University of Edinburgh. Unpublished.
- Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17:143–154.
- Wittgenstein, L. (1958). *Philosophical Investigations*. Blackwell.

## Appendix A

Stimulus word	LogCD	Mean AoA	Stimulus word	LogCD	Mean AoA
body	4.77087	3.833	history	4.09813	8.750
child	5.98587	4.333	knowledge	3.58889	9.500
door	6.38561	2.333	method	3.83751	10.750
hall	4.65299	4.667	month	6.83412	6.167
meat	4.20579	3.833	truth	4.30552	6.333
newspaper	3.83349	4.833	bright	3.94439	4.833
officer	5.45361	7.750	every	6.99156	6.000
sea	4.19893	4.000	foreign	4.26937	9.833
woman	5.83049	4.000	fresh	4.24607	6.500
car	6.25314	2.000	happy	5.95588	3.333
city	5.63426	5.167	important	6.61017	6.500
lake	2.89140	5.833	rich	5.07413	6.500
mountain	3.63197	4.667	soft	3.74442	3.833
river	3.77832	4.000	various	4.94526	9.917
valley	2.75932	7.333	different	6.41056	5.833
direction	4.19933	8.167	famous	3.58958	7.500
effort	4.28981	8.000	former	3.29353	11.333
hour	6.78776	6.417	golden	3.18198	5.833
law	4.47007	8.667	heavy	4.73867	4.500
life	5.41430	6.167	simple	5.08682	6.833
moment	4.79665	7.833	sudden	2.57531	7.167
opinion	3.84712	10.083	accept	4.47453	7.000
thought	4.24945	7.167	believe	4.68045	5.667
duty	3.74104	9.167	continue	4.24523	8.167
fact	5.41029	8.833	eat	5.45427	2.000
health	6.18103	7.167	grow	4.42738	4.167
include	4.13984	8.500			

## **Centre for Cognitive Science Publications**

Other Cognitive Science Research Papers, Working Paper collections and theses are available electronically at

<http://www.cogsci.ed.ac.uk/ccs/publications/>

For paper copies of the Centre's internal publications, please contact:

The Librarian  
Centre for Cognitive Science  
The University of Edinburgh  
2 Buccleuch Place  
Edinburgh EH8 9LW  
UK  
Email: [librarian@cogsci.ed.ac.uk](mailto:librarian@cogsci.ed.ac.uk)  
Fax: +44 131 650 4587