Title
Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Authors
Monaghan, Padraic
Charter, Nick
Christiansen, Morten H.

Publication Date
2003

Peer reviewed

Inequality between the classes: Phonological and distributional typicality as predictors of lexical processing

Padraic Monaghan (Padraic.Monaghan@warwick.ac.uk)
Nick Chater (N.Chater@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry, CV4 7AL, UK

Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology, Cornell University
Ithaca, NY 14853, USA

Abstract

Information about the syntactic category of a word can be derived from a number of complementary sources. We focus here on phonological and distributional cues for distinguishing nouns and verbs that have been proposed as useful for language acquisition. In this paper we assessed the extent to which this information affects lexical processing in adults. We hypothesised that the phonological or distributional typicality of a word with respect to its syntactic class would influence lexical access – words that were more typical of their class would be accessed more quickly. We tested this in three tasks: naming, lexical decision, and a noun/verb decision task. Words that were phonologically typical of their syntactic category were responded to more quickly in lexical decision and naming tasks. Distributional typicality related only to the noun/verb decision task.

Introduction

Information from a variety of sources has been hypothesised as useful for determining the syntactic category of a word. Such syntactic categorisation is important in language acquisition as a precursor to the child understanding the relationship between speech sounds and objects in the environment. In this paper we concentrate on phonological and distributional information that distinguishes nouns and verbs. Such information is useful for language acquisition, and early acquisition has been shown to exert a profound influence on processing in adults (Ellis & Lambon Ralph, 2000), therefore this information may impact on the processes of lexical access in adults. We present data from three lexical processing tasks: naming, lexical decision, and a noun/verb decision task to test this hypothesis. Before we discuss the phonological and distributional cues, we provide the background to this research in terms of syntactic categorisation in language acquisition.

There are many cues that are potentially useful in determining the syntactic category of a word, and we are interested in determining the identity, availability and use of these cues in early language acquisition. Cues may be derived from numerous sources – phonological, contextual, gestural, emotional, and so on – but cues are unreliable when considered alone for determining the syntactic (or semantic) category to which the word belongs. It may, therefore, be useful for the child to integrate multiple cues in language acquisition (Christiansen & Dale, 2001). This contrasts with a perspective that assumes that cues may act individually as triggers for structuring the language (Pinker, 1997).

Phonological and distributional cues are an interesting test case in exploring the issue of whether cues operate individually or integratively. As a test case, we assess the noun/verb distinction. In the next two sections we consider phonological and distributional cues that have been proposed as useful for syntactic categorisation. We then discuss how the "typicality" of a word with respect to these cues may influence lexical access. Finally, we present the studies testing phonological and distributional typicality in the lexical decision, naming, and noun/verb decision tasks.

Phonological cues for categorisation

Kelly (1992) reported several phonological variables that distinguished nouns and verbs. In English, for example, nouns tend to be longer than verbs, and tend to have stressed initial syllables as opposed to stress on the second syllable (e.g., ***re**cord* and *re**cord***). Such cues have been shown to predict category selection for nonwords, such that longer nonwords are more likely to be used in noun contexts by participants, and shorter nonwords used in verb contexts (Cassidy & Kelly, 1991). We compiled a set of 15 phonological cues that have been proposed as potentially useful in distinguishing words from different parts of speech (Durieux & Gillis, 2001; Kelly, 1992; Shi, Morgan & Allopena, 1998).

We derived all monosyllabic nouns and verbs that belonged to only one syntactic category, according to the CELEX database. This resulted in a set of 2029 nouns and 1986 verbs. We entered the 15 phonological cues as factors into a linear discriminant analysis with noun/verb status as dependent variable. Whereas each cue was unreliable for discriminating nouns and verbs when used alone, the combined set of cues correctly classified 60.9% of nouns and 67.0% of verbs in a leave-one-out cross-validation analysis.

We hypothesised that the extent to which a word shares phonological cues with other words in the same syntactic

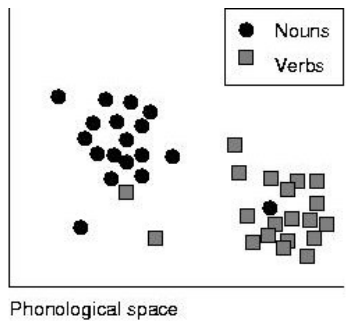

Figure 1. Idealised clustering of nouns and verbs in phonological space. Some atypical nouns are closer to the centre of the verb cluster than they are to the noun cluster.

category to which it belongs would influence the response latency to that word. Figure 1 shows the rationale for our hypothesis. Nouns tend to cluster in phonological space, as do verbs. However, some nouns are more like other verbs than other nouns in terms of their phonological cues. If phonological cues are used for acquisition of category information and this imparts an influence on the structure of the lexicon, then nouns close to the centre of the category will be responded to quickly. In contrast, "verby" nouns, with a clash between category and phonological form, will be responded to less quickly.

We assessed the probability of the category membership for each of the nouns and the verbs in the discriminant analysis. A noun that shares few phonological cues with other nouns has a low probability of noun membership, and a noun that shares many phonological cues has a high score. We predicted that nouns with high scores would be processed more quickly. We term this measure *noun/verb phonological typicality*. We next consider distributional cues for syntactic categorisation.

Distributional cues for categorisation

The contextual information of a word is extremely useful as a reflection of syntactic category. Indeed, the *specific* context in which a particular lexical token is found will generally provide sufficient information for the syntactic category to be guessed correctly. However, measures of distributional information for words have largely concentrated on providing generic methods for reflecting syntactic or semantic groupings. Such models do not assume that the input has been parsed, but rather have sought to provide simple n-gram models[1] of the language that are appropriate for the early stages of language acquisition. Such models rather drastically underestimate the potential influence of context, but are useful for providing a lower-bound for the availability of information from such a source.

Fries (1952) suggested that nouns and verbs tend to occur in different context templates. For example, only a noun occurs in the context 'the ____ is/was/are good'. To test the extent to which this information was useful, Redington, Chater and Finch (1998) generated cluster analyses of words according to counts of their contextual occurrence in a four-word window from a child-directed speech corpus, and found that the different syntactic classes clustered together with a high degree of accuracy and completeness.

Chiarello, Shears and Lund (1999) derived a measure of noun-verb distributional typicality. A co-occurrence context vector (from Lund & Burgess, 1996) was determined for each word. Next, the distance from the target word vector to that of each noun vector was computed, and subtracted from the mean distance from the target word vector to each verb vector. Nouns that occur in contexts similar to other nouns will have a higher score, whereas nouns which occur in atypical noun contexts will have a low score. This score was termed the noun-verb distance difference (NVDD). Hereon, we refer to this measure as *noun/verb distributional typicality*.

The overlap between the words with distributional typicality scores from Chiarello et al. (1999) and those from our own database was small, and so we reproduced distributional typicality scores using distances between context vector representations (McDonald & Shillcock, 2001). On the words that did overlap, our distributional typicality scores were correlated with those from Chiarello et al. (1999), $r(319) = 0.328$, $p < .001$. The correlation was low but highly significant – and, perhaps more importantly, our typicality measure was derived from co-occurrence vectors that produced a qualitatively similar level of semantic clustering to that of Lund and Burgess (1996).

We next present analyses of the role of phonological typicality in terms of phonological and distributional information for the processing of nouns and verbs. We assessed the contribution of phonological and distributional typicality in accounting for variance in response times in a naming task, a lexical decision task, and a noun/verb decision task.

Analysis 1: Noun/verb naming

Several variables contribute towards response times in a naming task. We wanted to assess the contribution of typicality over and above these other variables. Hierarchical regression analysis enables key variables to be entered after other variables have been accounted for as predictors of response times. In this regard, we employ the rationale of Spieler and Balota (1997), who found that several variables contributed towards accounting for variance in naming latencies. Position and manner of articulation of onset phoneme, log-frequency, orthographic neighbourhood size and length were all significant. Kacinik, Shears and Chiarello (2000) found that familiarity was also a significant contributor towards accounting for variance in lexical decision response times, and so we also included a measure of familiarity from Balota, Cortese and Pilotti (1999). The naming latency data were taken from Spieler and Balota's (1997) naming latencies for 2820 words, which were produced by 31 undergraduates at Washington University.

All unambiguous nouns and verbs with familiarity,

[1] N-gram models reflect the statistics of the language by counting co-occurrences in groups of *n* words. A bigram model will count the co-occurrences of two words, whereas a trigram model will count occurrences of triples of words.

Table 1. Results of hierarchical regression analyses of data from the naming task in Analysis 1. For nouns N = 714. For verbs N = 252.

NOUNS	***β* weight**	***t* value**	R^2
Step 1			.287
Onset-phoneme			
Step 2			.475
Log frequency	-.203	-3.458***	
Neighbourhood size	-.214	-4.307***	
Length	.051	1.047	
Familiarity	-.159	-2.683**	
Imageability	-.097	-2.300*	
Step 3			.479
Phonological typicality	.032	.699	
Distributional typicality	.959	-1.333	
VERBS	***β* weight**	***t* value**	R^2
Step 1			.556
Onset-phoneme			
Step 2			.654
Log frequency	.344	1.853	
Neighbourhood size	.125	1.081	
Length	.300	2.277*	
Familiarity	-.530	-2.883**	
Imageability	.108	1.132	
Step 3			.692
Phonological typicality	.241	2.465*	
Distributional typicality	.024	.240	

*p < .05, ** p < .01, *** p < .001

phonological and distributional typicality measures, and naming latency data were entered into a hierarchical regression analysis. In accordance with Spieler and Balota (1997), we first entered onset-phoneme characteristics in terms of the 13 phonemic features that they employed in their regression analyses of naming time (such as affricative, bilabial, stop, voiced). Next, we entered log-frequency, neighbourhood size, length and familiarity. Finally, we entered phonological and distributional typicality. The results are shown in Table 1.

The noun data accorded with the analyses on the whole set of words presented in Spieler and Balota (1997), with log-frequency and neighbourhood size as significant predictors. Familiarity was also found to be a significant predictor. For the verb data, neighbourhood size and log-frequency were not significant predictors, but length and familiarity were. Distributional typicality did not significantly predict naming response times, but phonological typicality did contribute towards accounting for variance in verb naming times. Phonological typicality acted in the predicted direction: *verbs that were more typical of the phonological properties of verbs were responded to more quickly than those that were less typical*. Words from different syntactic classes are influenced differentially by the psycholinguistic variables.

In order to test the maximal contribution towards accounting for variance by the typicality measures, we reversed the step order for the regression analysis. When the typicality measures were entered first, phonological typicality contributed significantly to accounting for variance for nouns, but not for verbs (β weight = -.164, t = -3.195, *p* = .002; β weight = .086, t = .714, *p* = .478, respectively). For distributional typicality, there was no evident contribution for noun naming latencies, but a marginally significant contribution to verb response times (β weight = -.076, t = -1.484, *p* = .139; β weight = .215, t = 1.788, *p* = .078, respectively). One of the phonological cues related to length of words in phonemes, and so overlap between phonological typicality and length may have obscured typicality effects in the initial hierarchical regression analyses.

Table 2. Results of hierarchical regression analyses of data from the lexical decision task in Analysis 2. N = 734 for nouns and N = 263 for verbs.

NOUNS	***β* weight**	***t* value**	R^2
Step 1			.211
Onset-phoneme			
Step 2			.516
Log frequency	-.282	-5.118***	
Neighbourhood size	-.002	-.034	
Length	.021	.461	
Familiarity	-.377	-6.788***	
Imageability	-.184	-4.613***	
Step 3			.522
Phonological typicality	-.079	-2.044*	
Distributional typicality	.007	-.199	
VERBS	***β* weight**	***t* value**	R^2
Step 1			.141
Onset-phoneme			
Step 2			.612
Log frequency	.207	1.054	
Neighbourhood size	.183	1.486	
Length	.135	.965	
Familiarity	-.919	-4.714***	
Imageability	-.088	-.872	
Step 3			.689
Phonological typicality	.339	3.459***	
Distributional typicality	.155	1.514	

p* < .05, ** *p* < .01, * *p* < .001

Analysis 2: Noun/verb Lexical Decision

Lexical decision latencies were taken from the corpus produced by Balota, Cortese and Pilotti (1999), with data from 30 Washington University undergraduate students who were tested on 2906 monosyllabic words, and an equal number of nonwords.

We performed a hierarchical regression analysis on the word lexical decision response times, with the 13 onset phoneme variables entered at the first step, log-frequency, neighbourhood size, length, and familiarity entered at the second step, and phonological and distributional typicality entered at the third step. The results are shown in Table 2.

The onset-phoneme variables did not account for as much variance in response times to lexical decision as they did in the naming task. At step 2, for nouns log-frequency, familiarity and imageability were significant predictors, whereas for verbs familiarity was the only significant predictor. Again, distributional typicality was not a significant predictor, though phonological typicality did predict response times for both nouns and verbs. In each case, *greater phonological typicality of the phonological properties of the word in its class reflected quicker responses to the word*.

When the step order was reversed in the regression analysis, phonological typicality was again a significant predictor of both noun and verb lexical decision latencies (β weight = -.134, t = -2.656, p = .008; β weight = .251, t = 2.102, p = .039, respectively). Again, distributional typicality was not found to contribute towards accounting for variance in response times (for nouns, β weight = -.073, t = -1.440, p = .151; for verbs, β weight = .024, t = .203, p = .840).

Experiment 1: Noun/verb Decision Task

Kacinik and Chiarello (2002) found that noun-verb distributional typicality was a significant contributor to accounting for the variance of response times in a noun/verb decision task. In a regression analysis, distributional typicality and imageability were significant for predicting response times for both nouns and verbs, with familiarity also contributing to verb response time. We wished to test whether both phonological and distributional typicality contributed towards accounting for response times.

Method

Participants Forty native English speaking postgraduates and undergraduates at the University of Warwick participated, and were paid £1 for taking part.

Stimuli We used all 71 verbs for which we had measures of familiarity and phonological and distributional typicality. We paired these with 71 nouns which were selected to match the range of frequency and length found in the verb stimuli.

Procedure Participants were required to judge whether a word was a noun (press "n") or a verb (press "v"). For each trial, a fixation cross appeared at the centre of the screen for 400ms, followed by the 100ms blank screen, followed by the word, which disappeared on response and was replaced after 1000ms with the fixation cross for the next trial. The 142 trials were preceded by 8 practice trials comprised of 4 nouns and 4 verbs that were not in the stimulus set. As in Kacinik and Chiarello (2002), participants were instructed that nouns were words naming a quality, person, place or thing, and that verbs were words that refer to an action or the occurrence of an event, and examples were given of each.

Results

We performed a hierarchical multiple regression analysis on response times for nouns and verbs separately, with onset phoneme variables entered at the first step, log-frequency, neighbourhood size, length, and familiarity entered at the second step, and phonological and distributional typicality

Table 3. Results of hierarchical regression analyses of data from the noun/verb decision task in Experiment 1. N = 71 for both nouns and verbs.

NOUNS	*β* weight	*t* value	R^2
Step 1			.523
Onset-phoneme			
Step 2			.598
Log frequency	.052	.333	
Neighbourhood size	.238	1.810	
Length	.285	2.421*	
Familiarity	-.099	-.674	
Imageability	-.642	-1.705	
Step 3			.698
Phonological typicality	.048	.501	
Distributional typicality	-.489	-4.034***	
VERBS	***β* weight**	***t* value**	R^2
Step 1			.464
Onset-phoneme			
Step 2			.530
Log frequency	-.089	-.487	
Neighbourhood size	.159	1.283	
Length	.251	2.018*	
Familiarity	-.150	-.782	
Imageability	-.214	-.631	
Step 3			.548
Phonological typicality	-.029	-.305	
Distributional typicality	-.182	-1.902	

*p < .05, ** p < .01, *** p < .001

entered at the third step. Table 3 shows the results for nouns and for verbs.

Both nouns and verbs had length as a significant predictor. Neither nouns nor verbs demonstrated an effect of phonological typicality, though we replicated the effect of noun-verb distributional typicality for nouns found by Kacinik and Chiarello (2002). Nouns that occurred in more typical contexts were responded to more quickly. However, there was a marginally significant effect of typicality for verbs, which was in the same direction – meaning that verbs which occurred in more typical noun contexts were responded to more quickly. This is in contrast to the finding by Kacinik and Chiarello (2002), who found that verbs which were more distributionally typical of the verb class were responded to more quickly. Kacinik and Chiarello (2002) used a larger set of words (152 nouns and 137 verbs) and so the additional power of their analyses may have revealed effects that were not significant in our data. Another possibility is that the distributional typicality measure we employed was somehow importantly different to that used in their study. The correlation between the Chiarello et al. (1999) measure and our distributional typicality measure accounted for only a moderate amount of variance.

Local distributional typicality

Distributional typicality, as measured by NVDD, was not successful in predicting response times for naming or lexical decision tasks. Our measure of distributional typicality was derived from co-occurrence counts across a window of 10 words, taking into account only content words. Such broad windows of co-occurrences blur the syntactic information, and have been used rather as a reflection of semantics. Measures of contextual typicality in terms of co-occurrence with function words, and in a more local contextual window may better reflect the syntactic typicality of nouns and verbs.

To this end, we counted the proportion of times that each unambiguous noun and verb occurred immediately after a set of context words in the 100 million word British National Corpus (Burnard, 1995). We chose the twenty most frequent words from the corpus as the context words. This provided a 20-dimensional vector for each word, which we entered into a discriminant analysis with noun/verb as dependent variable. As for the phonological typicality measure, we derived a probability of classification as a noun or a verb for each word. The classification was extremely accurate, with 99.0% of nouns and 88.4% of verbs correctly classified in a leave-one-out cross validation analysis. We repeated the multiple regression analyses for naming response time and lexical decision response time, replacing noun/verb distributional difference with the new measure of local distributional typicality. The results were very similar, with phonological typicality significant for the same tests, and distributional typicality not contributing significantly to accounting for variance in latencies.

Discussion

We found a significant effect of phonological typicality on verb naming latencies. We also found a significant effect of phonological typicality on lexical decision latencies for both nouns and verbs. This latter result is especially surprising, as speech production is not directly implicated in lexical decision, and yet the extent to which a word conformed to phonological typicality for its class influenced processing time. This effect may reflect the influence of graphemic typicality, as nouns share similar letters in similar positions within the word. It is still a surprising effect given that the onset variables account for so much less variance in the lexical decision task compared to the naming task.

The reliability of phonological typicality across naming and lexical decision tasks for verbs indicates that it is an important factor in describing the processing of single words. The extent to which a word resembles the phonological characteristics of other members of its syntactic class influences reaction times. Words with a good match are responded to more quickly than words where there is a mismatch between category and typicality. This suggests that syntactic information is engaged during lexical access in single-word reading tasks. We hypothesise that the speeded access for lexical items typical of their syntactic category is due to cohort activity from other items in that category. This is reminiscent of the gang-effect explanation proposed by Kelly, Shillcock and Monaghan (1996) for activating the class of function words. Function words are closer to one another in phonological space than they are to content words, and activation of a target function word is assisted by partial activation of other function words that are close in phonological space. The analyses presented here indicate that further discriminations can be made *within* the content word class, and, furthermore, that the gang-effect is greatest for words that are closest to the phonological exemplar for their category.

Cassidy and Kelly (1991) have focused primarily on individual phonological cues, such as syllable length. Our analyses have merged several phonological measures together, though it is possible that certain phonological cues may have an individual impact on determining phonological typicality. We performed a stepwise discriminant analysis on noun/verb category for all unambiguous monosyllabic words with the 15 phonological cues entered as factors. Factors were entered into the analysis if the probability of F was less than .05. Five cues were entered: length in phonemes (P1), proportion of coronal consonants (P2), final-consonant voicing (P3), number of consonants in the onset (P4), and proportion of nasal consonants (P5). The leave-one-out classification was similar in accuracy to the original discriminant analysis we performed (61.8% of nouns and 67.3% of verbs correctly classified).

We entered these five phonological cues in place of phonological typicality in the third step of the regression analysis. For noun naming, only P3 was a significant contributor. For verb naming, P1, P2, P4 and P3 all contributed. No cue contributed for noun lexical decision, and for verb lexical decision, P1, P3 and P5 contributed. P3 contributed as a predictor for three measures, but no individual cue reflected the effect of phonological typicality found for noun lexical decision. In each analysis, different cues seemed to contribute towards accounting for variance, and this suggests an integrated use of the various cues better accounts for the naming and lexical decision latency data. Integrated cues provide a better explanation of the typicality effects than do single trigger cues.

We have shown that differences between nouns and verbs are evident even without the oft-cited difference in terms of syllable length as a single cue all the words in our analyses were monosyllabic. We predict that the effects we have found for monosyllabic words will be stronger still for polysyllabic words. Discriminant analysis conducted on unambiguous nouns/verbs of all syllabic lengths resulted in a more accurate classification than just on monosyllabic words (approaching 75% accurate classification). For polysyllabic words, phonological typicality matches syntactic class better, and we suggest that mismatches between phonological properties and syntactic class would impact on processing to a greater degree.

Distributional typicality was not found to predict naming or lexical decision response times. This corroborates Kacinik and Chiarello s (2002) study where distributional typicality was only a predictor for a noun-verb decision task. We partially reproduced the effect of distributional typicality on the noun-verb decision task, when the contextual window was wide (+/-5 words), or narrow (-1 word). The noun-verb decision requires the participant to evoke implicit knowledge of words that is difficult to retrieve. Distributional typicality appears to assist the elicitation of such information, whereas phonological typicality, in contrast, did not predict noun/verb decision times. Noun/verb decision, therefore, appears to be tapping rather different processes to those involved in naming and lexical decision tasks. These latter tasks have been taken to reflect general processes of lexical access.

We have been successful in demonstrating that nouns and verbs have different variables as predictors. The contribution of different variables in predicting variance in naming response times indicates that even grapheme to phoneme correspondence is modulated by syntactic category. For instance, log-frequency is predictive only of naming time for nouns, and, indeed for verbs, it has a non-significant but reversed effect. Consequently, large scale item analyses of naming or lexical decision response times ought to attend to the possibility of varying factors involved in lexical access for different syntactic categories.

Conclusion

We found no effects of distributional information for naming and lexical decision tasks. The influence of such cues are more marked for the noun/verb decision task. However, we have shown that phonological typicality influences naming and lexical decision response latencies in adult readers. We have focused on phonological cues that have been proposed to be involved in constructing syntactic category information in early language acquisition. The influence of these phonological cues appears to be integrative: Single cues are ineffective in accurately discriminating between nouns and verbs, though accuracy is significantly greater than chance when several cues are entered simultaneously into a discriminant analysis. In short, phonological cues employed in syntactic categorisation exert an influence on lexical access, similar to the influence of age of acquisition for single word processing. The processes of early language acquisition can be observed in adult language processing.

Acknowledgments

This research was supported by a Human Frontiers of Science Program grant. Thanks to Christine Chiarello for providing an electronic database of NVDD scores.

References

Balota, D.A., Cortese, M.J. & Pilotti, M. (1999). Item-level analyses of lexical decision performance: Results from a mega-study. In *Abstracts of the 40th Annual Meeting of the Psychonomics Society* (p. 44). Los Angeles, CA: Psychonomic Society.

Burnard, L. (1995). *Users Guide for the British National Corpus.* British National Corpus Consortium, Oxford University Computing Service.

Cassidy, K.W. & Kelly, M.H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, **30**, 348-369.

Chiarello, C., Shears, C. & Lund, K. (1999). Imageability and distributional typicality measures of nouns and verbs in contemporary English. *Behavior Research Methods, Instruments, and Computers*, **31**, 603-637.

Christiansen, M.H. & Dale, R.A.C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, pp..

Durieux, G. & Gillis, S. (2001). Predicting grammatical classes from phonological cues: An empirical test. In J. Weissenborn & B. Höhle (Eds.) *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language Acquisition* Volume 1. Amsterdam: John Benjamins.

Ellis, A.W., & Lambon Ralph, M.A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26**, 1103-1123.

Fries, C.C. (1952). *The Structure of English: An Introduction to the Construction of English Sentences.* New York: Harcourt, Brace & Co.

Kacinik, N. & Chiarello, C. (2002). Predicting noun and verb latencies: Influential variables and task effects. *Proceedings of the 24th Cognitive Science Society Conference.* Mahwah, NJ: Lawrence Erlbaum Associates, pp.524-529.

Kacinik, N., Shears, C. & Chiarello, C. (2000). Familiarity for nouns and verbs: Not the same as, and better than, frequency. *Proceedings of the 22nd Cognitive Science Society Conference*. Mahwah, NJ: Lawrence Erlbaum.

Kelly, M.H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, **99**, 349-364.

Lund, K. & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, **28**, 203-208.

McDonald, S.A. & Shillcock, R.C. (2001). Rethinking the Word Frequency Effect: The neglected role of distributional information in lexical processing. *Language and Speech*, **44**, 295-323.

Pinker, S. (1997). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**, 425-469.

Shi, R., Morgan, J. & Allopena, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*, **25**, 169-201.

Shillcock, R.C., Kelly, M.L. & Monaghan, P. (1996). Modelling within-category function word errors in language impairment. In W. Ziegler & K. Deger (Eds.). *Clinical Linguistics and Phonetics*. London: Whurr.

Spieler, D.H. & Balota, D.A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, **8**, 411-41.