Second Language Research 11, 2 (1995); pp. 121–136

Vocabulary skill: single-case assessment of automaticity of word recognition in a timed lexical decision task

Norman Segalowitz *Concordia University,*
Vivien Watson *Concordia University* and
Sidney Segalowitz *Brock University*

This study illustrates, in the context of vocabulary assessment research, a procedure for analysing a single subject's variability of response times (RTs) in a simple, timed lexical decision task. Following the interpretation developed in Segalowitz and Segalowitz (1993) for RT variability as reflection of the automatic/controlled nature of underlying processing mechanisms, it was possible to draw conclusions about the extent to which second language English word recognition in this subject was subserved by automatic as opposed to controlled processes. The study also examined the development of automaticity in word recognition skill for a small, selected vocabulary as a function of reading experience during a three-week testing period. The general implications of this methodology for assessing vocabulary skill in a single case are discussed.

Most research on vocabulary development has focused on the content of word knowledge, that is, what and how many meanings does a person know, how sophisticated is this knowledge of meaning and what is the format of a word's mental representation (Colley, 1987; Cooksey and Freebody, 1987; Graves *et al.*, 1987; Meara and Buxton, 1987). In this article we focus on a different but related aspect of vocabulary skill, one that has to do with the nature of the underlying *process*, specifically, the assessment of the degree of automaticity of word recognition and how to measure this in a single individual. We show how reaction times (RTs) taken from a simple, timed lexical decision task can be used for this purpose. The advantage of this task is that it is relatively easy for the researcher to administer, and is straightforward for the subject to perform. The subject simply views a string of letters presented on a computer

The authors acknowledge support for this research from a Quebec Ministry of Education grant (FCAR) to the first author. The authors thank Elizabeth Gatbonton and Mel Komoda for their thoughtful comments on earlier drafts of this article.

Address for correspondence: Norman Segalowitz, Department of Psychology, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec H4B 1RS, Canada.

screen and presses a key or button to indicate whether the string forms a valid word (see, for example, den Heyer *et al.*, 1988; Henderson, 1982; Taft, 1991, for a discussion of the issues involved in this technique). The principal focus of this study is to demonstrate the practicality – and some of the limitations – of a particular method of assessing word recognition automaticity in a single individual.

There are a number of reasons for interest in the question of whether and to what degree word recognition has been automatized. Obviously, word recognition is fundamental to reading and listening comprehension for the simple reason that language messages are composed of strings of words. Beyond this, there is the very important issue about how the language comprehender is deploying his or her psychological resources when comprehending a message. If an individual's mastery of vocabulary is weak in the sense that too many attentional resources have to be assigned to the recognition of the individual words in the incoming message, then there will likely not be sufficient resources available to integrate the new information with older knowledge, or to develop a representation of the text or story as a whole. Thus a full picture of vocabulary skill development should include consideration of the development of underlying word recognition mechanisms. Recent work in both first and second language reading underscores the importance of this issue. For example, Graesser *et al.* (1980) found that slower readers were distinguished from faster readers not in the speed at which they globally integrate information across a text but in the speed at which they processed local information at the level of individual words or phrases. Favreau and Segalowitz (1983) found that bilinguals who read their second language significantly more slowly than their first, but who were otherwise highly fluent in both languages, exhibited less automaticity in word recognition in the slower language (see also Segalowitz, 1986; 1991). Evidence of this type suggests that one goal of vocabulary development, in addition to increasing the size of one's vocabulary, should be the acquisition of highly efficient processes for accessing meanings already in one's vocabulary. To study this aspect of vocabulary development, researchers will need tools to permit 1) the evaluation of word recognition efficiency; and 2) changes in the way an individual exercises this skill as a function of practice or learning experience. This article presents a method for conducting just such an assessment.

The main theoretical underpinnings for the approach used in this study are the following (see also Segalowitz and Segalowitz, 1993):

1) *Word recognition is complex.* It is assumed that word recognition or lexical access is a relatively complex activity involving a number of component processes (see Forster, 1989, for a discussion of theoretical issues involved here). For example, the full process of understanding a word during reading may, depending on the context and the reader's level of skill, involve visual analysis of the printed stimulus (Besner and Johnston, 1989), the generation of expectations derived from the words orthographic redundancy (Favreau *et al.*, 1980), the generation of a phonological code based on the graphemic properties of the word (Perfetti and McCutchen, 1982; Segalowitz and Hébert, 1990), and possibly some evaluation procedure that selects and confirms a choice made from among competing candidate meanings. If the word is presented as part of a running text, then higher-level analyses of the text itself may play a role in identifying the meaning of the target word (Stanovich, 1980). Each of these processes can, in turn, be decomposed into components. For example, information may be retrieved from long-term memory – information about letter shapes, phonetic values or semantic features, depending on the process concerned – and held in working memory for some kind of local search and analysis (Perfetti, 1985).
2) *There is significant serial organization of the components underlying word recognition.* Word recognition is likely to involve a cascading organization of the underlying component processes, with the possibility of some parallel processing, overlap in processing times of components and interactions between some of these components (Perfetti, 1985). For example, processes concerned with visual analysis of letter shapes and those concerned with the knowledge about the language's orthographic redundancies may interact. Nevertheless, word recognition clearly does take some finite amount of time and we focus here on the rate-determining and mechanisms operating sequentially that affect how long word recognition will take (Segalowitz and Segalowitz, 1993).
3) *Skilled word recognition involves a blend of automatic and controlled processes.* Every skill, especially a complex cognitive skill such as lexical access, depends on a variety of underlying mechanisms for its smooth and accurate execution (e.g., those described above). Any given component mechanism can be said to be highly automatized in the following sense. It makes its contribution in a ballistic fashion – once activated it continues until completion and cannot be stopped. It is informationally encapsulated (Fodor, 1983; Stanovich, 1991), that is, it draws little or no processing resources away from other ongoing activities and

> cannot be interfered with by other mechanisms. Finally, because the mechanism operates without interference from other sources of information, it generally operates very fast and is relatively stable in execution time. In contrast, other component mechanisms can be said to be relatively more controlled. They involve decision-making and evaluation of information, and hence respond to, perhaps even require, information from other sources in order to complete their contribution. As a result, they generally operate slowly and are relatively more variable in their execution time. Thus the constellation of component operations underlying skilled performance will involve a blend of automatic and controlled components (Jacoby, 1991). The more highly skilled the performance, the more this blend will be characterized by the automatization of those components capable of being automatized. (It is important to remember, however, that there will always remain controlled components in any skilled performance, and that these contribute just as significantly as do the automatized components to rendering the performance 'skilled'.)

When word recognition – or, for that matter, any complex skill – is viewed this way, it is possible to analyse RTs and their variability to shed light on the relative balance of automatic and controlled processes that make up the blend of mechanisms underlying performance. Total response time (RT) will be affected by the way underlying components contribute to overall performance. Each of the slower components – typically, controlled processes, associated with decision-making – will add considerably to the overall RT and to the variability of the overall RT. Each of the faster components – typically, the automatized processes, associated with modularized pattern analysis mechanisms – will add a smaller increment to the overall RT and to its variability. Thus, if we have two samples of performance – one faster and more stable (in timing) than the other – we can ask the following question: Does the observed difference in RT and variability warrant the conclusion that the faster performance reflects the operation of a more highly automatized blend of underlying processing components?

A single answer to this question is not possible. Faster performance may, for example, reflect a general speed-up of all underlying mechanisms, not a change in the blend of automatic and controlled components. Thus, for example, if the individual has gained a great deal of practice in word recognition between time 1 and time 2, we could expect word recognition RT to be speeded up. As noted above, RT can be viewed as being significantly determined in an additive fashion by the processing times of the underlying components, and hence the variability of the overall RT will be determined additively by the variances of the underlying component reaction times. If at time 2 RT is half what it was at time 1, then speed-up effects alone predict a corresponding proportional change in the overall SD (i.e., SD will be reduced by half). Thus simply observing a change in RT and SD does not warrant the conclusion that there has been a change in the blend of underlying mechanisms.

Faster performance can, nevertheless, sometimes reflect a change in the blend of underlying mechanisms. For example, some of the mechanisms that are relatively slow and quite variable in their times of execution at time 1 might either drop out entirely or become modularized and operate more ballistically at time 2 – that is, faster and in a more stable fashion. This time, reduction in response time will be accompanied by a more than correspondingly proportional reduction in variability. Thus to warrant the conclusion that there has been a change in the blend of underlying mechanisms – and not just a speed-up effect – there needs to be a reduction in SD that is more than proportional to the reduction in RT.

The coefficient of variation (CV), defined as the SD/RT, provides a conceptually useful index for this purpose since it reflects the SD *as a function of a given level of RT* (Segalowitz and Segalowitz, 1993; Watson, 1993; see also Herrington *et al.*, 1994, for a related discussion). For purposes of statistical analysis, an alternative but equivalent way of analysing changes of variability is to adjust proportionally all scores so that the RT mean is equal across the conditions to be compared. The resulting SDs will be correspondingly changed proportionally. Significant decreases in these adjusted SDs will now reflect change that is more than proportional to changes in RT, and will support the conclusion that increased automaticity, and not just speed-up, has occurred. Details of the statistical procedures involved here are given in the analyses section below.

In our study, a single subject with moderate to high-level reading skill in English was given a lexical decision task on four equally spaced occasions across a span of three weeks. The task was designed to assess two aspects of his vocabulary skill. The first concerned the automaticity in word recognition for words that occur with different levels of frequency in the language. Here we obtained a measure of the degree of automaticity for words that were known to be generally very high in their frequency of appearance in written text, and for words known to be significantly rarer in occurrence. The expectation here was that, unlike native speakers who would be expected to show high levels of automaticity in word recognition throughout a wide range of vocabulary, a second-language user of English with moderate skill would reveal such automaticity, if at all, in a more restricted fashion. Such a demonstration would increase our confidence in the analytical techniques used here and would justify their development for use in more complex situations. For example, one might wish to extend this method of analysis to comparison of selected domains within a vocabulary, to comparison of the effects of different language learning backgrounds, different motivations, teaching methods and so on.

The second aspect of this study concerned the acquisition of automaticity in word recognition as a function of 'extracurricular' reading experience during the three-week period of testing. Both in the second-language learning literature and in the cognitive psychological literature, there is growing consensus that optimal learning conditions are those in which the to-be-learnt material is processed 'deeply', that is, in terms of its core meaning (Craik and Lockhart, 1972; Hashtroudi, 1983). There is, of course, continuing discussion on the reasons for this, with references made to the role of elaboration, encoding specificity and depth of processing in enhancing memory (Tulving, 1979). Nevertheless, the picture that emerges from this research is that the conditions that most facilitate skilled recognition of words is repeated exposure to the words in a context that requires semantic processing, and where there is a consistent association between stimulus words and their meanings (Gatbonton and Segalowitz, 1988). For this reason, a task was selected requiring the subject to analyse a scientific article for a genuine, personal purpose not having to do with the study (namely, as part of his preparation for a research project in the next semester). His performance with selected words that appeared in the article became the focus of this study. In particular, we were interested to see if the focused reading experience would result in increased automatization of the recognition of these words compared to an appropriately chosen set of control words.

Comparison data from native speakers of English, tested for one session only, were also obtained to provide a point of reference for interpreting the data gathered from the principal subject.

I Method

1 Subjects

The principal subject was G.S., a 26-year-old male whose mother tongue is Greek and who speaks and understands English at a moderately strong level. His self-rated English language skills on a scale of 1–5 were 4, 4, 4 and 3 for speaking, listening, reading and writing respectively, where 1 = 'no ability at all' and 5 = 'native-like ability'. The student was enrolled in a qualifying year in psychology at Concordia University, an English language university in Montreal, having completed his BA in psychology in Athens. The subject was paid $20 for his participation.

For comparison purposes, three native speakers of English, aged 19–23, were also tested (NS1, NS2, NS3), in one session only, each with the stimuli corresponding to G.S.'s first session. All were university students. Subjects were pair $5 each for their participation.

2 *Materials and apparatus*

The main set of materials consisted of 1500 words and nonwords of 4–7 letters in length selected from a larger set of 4552 words (*n* = 970, 1092, 1269 and 1221 for each length respectively) according to the following criteria.

a Base words: A set of base words was selected from four frequency 'bands' of the language, based on the Kucera–Francis norms as given by the Oxford Psycholinguistic database (Quinlan, 1992). The lowest frequency band (band 1) contained words with a frequency of 5–10 words per million (*n* = 1482). The remaining bands had frequencies of 11–25 (band 2: *n* = 1176), 26–100, (band 3: *n* = 1260) and 101 or higher (band 4: *n* = 634). For each of the four test sessions, 30 different words were randomly selected without replacement from each of the four frequency bands.[1]

b Base nonwords: Orthographically regular pronounceable nonwords were created by changing one letter in a given base word. For each word and nonword one can calculate an *n*-value, the number of (other) real words that can be formed by changing a single letter of the original string (see e.g., Taft, 1991). For the purposes of this study, the wordlist against which each nonword string was compared was the list of 4552 Oxford database words. The resulting nonwords had, overall, an *n*-value equal to or greater than the real words. This procedure ensured that the pool of nonwords did not have a greater 'nonword' look about them than did the real words. For each for the four test sessions, 30 different nonwords were derived from words not being used as base words and randomly selected without replacement from each of the four frequency bands.

---

[1] Owing to a programming error, in session 1 the numbers of words in frequency bands 1–4 were 30, 40, 28 and 22 respectively instead of 30 in each.

c Studied words (SW) and control words (CW): There were 30 additional words selected for purposes of a separate analysis. Of these, 15 words were to be found in a research article that was studied intensively by the subject during the three-week period following the first session. The remaining 15 words, which did not occur in the article, served as control items. The same SW and CW stimuli appeared in each test session.

d Repeated nonwords: To keep the total number of words and nonwords presented to the subject equal, and additional 30 nonwords were added to the stimulus set and repeated in each session. Table 1 summarizes the characteristics of each of these stimulus sets.

Table 1 Frequency (occurrences per million) and *n*-value characteristics of stimuli

Stimulus category		Mean frequency	*n*-value
Base words			
Session 1	(*n* = 120)	75.8	1.89
Session 2	(*n* = 120)	105.2	2.21
Session 3	(*n* = 120)	80.0	1.36
Session 4	(*n* = 120)	76.1	1.81
Base nonwords			
Session 1	(*n* = 120)		2.19
Session 2	(*n* = 120)		2.51
Session 3	(*n* = 120)		2.48
Session 4	(*n* = 120)		2.50
Studied words	(*n* = 15)	63.4	0.33
Control words	(*n* = 15)	54.4	1.73
Repeated nonwords	(*n* = 30)		2.20

In each session, the 150 words and the 150 nonwords were presented singly in random order in the centre of the computer screen controlled by a Macintosh LC 475 running an experimental program written in Hypercard 2.2. The stimuli were printed in Palatino font 25 point and presented for up to 3 seconds or until the subject responded. The subject responded by pressing the space bar with his right (preferred) hand to indicate the item was a word, and a letter key in the row above the space bar with his left hand to indicate the item was not a word.

3 Procedure

Subject G.S., performed the lexical decision task on four occasions over a period of three weeks. Each session lasted about 20 minutes.

Between sessions, he met with the first author in a tutorial session to discuss assignments related to the assigned article. The native speakers of English performed only the first session.

4 Analyses

Two main analyses were conducted on the data obtained. The first aimed at determining the degree of automaticity of word recognition across the entire vocabulary, while the second aimed at determining the development of G.S. of automaticity of word recognition skill as a function of reading experience.

Before proceeding with a detailed presentation of the data analyses, it is necessary to explain some of the logic behind the statistical procedures employed. Evaluation of the automaticity hypothesis as we have framed it requires testing hypotheses not about the mean reaction times but about the degree of variation of those reaction times with respect to the mean reaction time. That is, the size of the subject's variance is the critical issue, but only with respect to his or her own average reaction time. This is why, as described earlier, the coefficient of variation is more appropriate than the standard deviation.

It turns out that direct statistical comparisons of two coefficients of variation, as we would find in a case study, is not a straightforward matter. There is no clear statistical distribution against which to compare them. We can avoid this problem by circumventing the need to use the coefficient of variation with the following two-step procedure. First we adjust the scores from each condition so that they are equated for average reaction time; secondly, we make a direct comparison of the variances associated with each condition. The first is done by simply dividing each RT in a given condition by the average for that condition to create a dataset that is relative to a unit RT. Because the original RTs come from a ratio scale, this adjustment will change the variances in a controlled way. If two sessions have the same degree of variability as indexed by the standard deviations of original RTs relative to the average RT in that session, then adjusting each individual data point in terms of a unit RT will produce standard deviations that are similar to each other.

The second step involves comparing the degree of variability. The traditional method is to perform the F-max test: finding the ratio of the larger variance to the smaller one and looking up the result on the F-table (with the degrees of freedom associated with the numerator and denominator). Unfortunately, this test is very highly sensitive to non-normality of data distribution, and reaction times are notoriously non-normally distributed. The resulting instability of the F-max test is that sometimes the .05 level of probability is exaggerated and sometimes it is underestimated.

Two alternatives exist, tests devised by Levene (1960) and by O'Brien (1981). The Levene test is based on the notion that the absolute deviation from the mean (or the median) will be smaller in a sample with a smaller variance (Howell, 1992). This test suffers, however, from relatively low power. O'Brien (1981) outlined a simple formula that transforms the raw data into a dataset that reflects the variance directly, where the average transformed score equals the sample's unbiased variance estimate. The formula[2] is straightforward to calculate, and will be illustrated in our case-study example.

II Results

Analyses were performed only on data from trials in which the stimulus was a word (base word, studied word or control word) and for which the response was correct. G.S.'s overall error rate was 5.0% on base words, 2.0% on studied words and 3.0% on control words across all four sessions. The error rates on base words for the three native speakers were 7.5%, 6.6% and 5.0%. Overall error rates on nonword trials were 5.0%, 5.0%, 0.0% and 2.5% for G.S. and the three native speakers respectively. NS1 made two errors on the 15 words that served as studied words for G.S.; there were no other errors to studied or control words by the native speakers.

Table 2 shows the changes in G.S.'s RTs to base words as a function of session and frequency band. Analysis of variance of these data revealed a significant main effect for session ($F(3,443) = 4.24$, $p < .007$) and a significant main effect for frequency band ($F(3,443) = 33.19$, $p < .001$). The interaction was not statistically significant ($F(9,443) < 1$).

Table 2 also shows the changes in CV to base words as a function of session and frequency band. As described earlier, these figures reflect the behaviour of the SD or RT, for a given level of RT. For purposes of analysing this adjusted measure of variability, the RTs were transformed as follows. First, the data in each of the four sessions were divided by the mean RT of the session to create a

---

[2] O'Brien's (1981) formula for transforming the data is

$$r = \frac{(n - 1.5)n(y - \bar{y})^2 - .5s^2(n - 1)}{(n - 1)(n - 2)}$$

where r is the transformation of data entry y, n = the size of the cell sample, y = the RT, y = the mean RT in cell, and s = SD. In our study, each RT entry was first transformed by dividing it by the cell mean RT to obtain a unit RT. These unit RTs then served as the data (the *ys*) for the transformation described above. See text for full explanation.

Table 2 Mean response times (RT) and coefficient of variation (CV) of response time for base words

	Frequency							
	Band 1		Band 2		Band 3		Band 4	
	RT	CV	RT	CV	RT	CV	RT	CV
Subject G.S.								
Session 1	772	.228	709	.197	655	.188	573	.155
Session 2	751	.332	698	.252	607	.226	585	.175
Session 3	722	.175	620	.153	594	.242	557	.140
Session 4	763	.192	751	.208	629	.242	592	.183
Native-speaking subjects (session 1 only)								
Subject NS1	615	.242	546	.190	570	.194	522	.182
Subject NS2	712	.202	645	.180	634	.252	621	.175
Subject NS3	664	.254	637	.237	587	.217	542	.205

dataset with equal, unit RTs. Next, the transformation recommended by O'Brien (1981) and described in footnote 2 was performed on these data. Finally, the transformed data were submitted to a 4 × 4 analysis of variance with the factors session and frequency band. The analysis revealed no significant main effect for session ($F(3,438) = 1.17$, $p > .05$) but there was a significant effect for frequency band ($F(3,438 = 3.43$, $p < .02$). The session by frequency band interaction was not significant ($F < 1$).

Table 2 also shows the analogous RT and CV data from the single session performed by the three native speakers of English. These analyses revealed that all three subjects showed a significant frequency band effect whereby RT became faster with the higher word-frequency bands (for subjects NS1, 2 and 3 respectively: $F(3,107) = 3.26$, $p < .03$; $F(3,116) = 3.43$, $p < .02$; $F(3,109) = 4.20$, $p < .01$). To analyse the variability of RT, the data was transformed as described above. Analysis of variance of the transformed data did not yield a significant frequency effect for any of the native speakers ($F(3,107) = 1.87$, $p > .10$; $F(3,116) = 1.52$, $p > .20$; $F(3,113) < 1$).

Table 3 shows G.S.'s RTs and CVs for studied words and control words as a function of session. A two (studied words, control words) by four (sessions) analysis of variance of the RT data revealed no significant main effects or interaction (all $Fs < 1.6$, $p > .20$).

For purposes of the variability analysis, the RTs were next appropriately transformed (dividing each RT by the cell mean and then transforming to O'Brien's r statistic). A 2 × 4 analysis of variance was again used with the transformed data. The analysis yielded no significant main effects or interaction. These transformed data, did, however, contain extreme outliers which could have affected

Table 3 Subject G.S.'s mean response times (RT) and coefficient of variation (CV) of response times for studied and control words

	Studied words		Control words	
	RT (msec)	CV	RT (msec)	CV
Session 1	619	.197	582	.162
Session 2	600	.145	572	.153
Session 3	546	**.125**	**561**	.189
Session 4	578	**.132**	**622**	.190

negatively the outcome of the analysis. The data were reanalysed, therefore, in the following way. First, words with a Kucera–Francis frequency greater than a 100 were removed from the analysis, on the assumption that learning is more likely to occur for less frequent items. This resulted in the number of studied words dropping to 10 and control words to 12. Next, the two highest and two lowest outliers were removed from SW and CW in each session. The resulting analysis of variance revealed a significant sessions effect (F(3,54) = 3.87, $p < .02$) and a significant word condition effect (F(1,54) = 5.77, $p = .02$). The interaction effect was not significant (F(3,54) = 1.14, $p > .05$). Inspection of the CV patterns for studied versus control words across sessions in Figure 1 suggests, nevertheless, that variability for studied words was reduced across sessions while for control words it was not. Indeed, when separate analyses were performed on SW and CW data, a significant main effect for session appeared for SW (F(3,22) = 3.56, $p < .05$) but not for CW (F(3,32) = 2.09. $p > .10$).

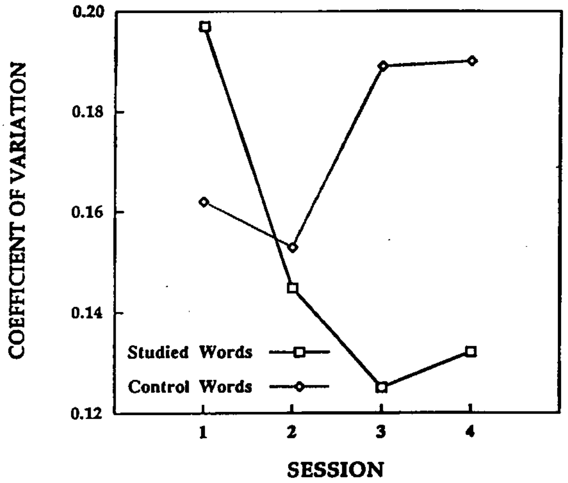

Figure 1 Changes in the coefficient of variation of G.S.'s response times for studied words and control words across four testing sessions

III Discussion

The goal of this study was to present a way of analysing data from a simple, timed lexical decision task to assess the degree of automaticity of word recognition skill in a single individual.

Data were looked at with respect to two specific issues. First, the study investigated the degree to which subject G.S. revealed his word recognition processes to be automatized. From Table 2 it can be seen that, as a function of increasing frequency of occurrence in the language, RT became faster and this was confirmed by analysis of the RT data. This could reflect a perceptual fluency effect whereby the more an item has been seen, the faster one is able to carry out the perceptual analyses required to identify it, which is a speed-up effect. As was pointed out in the introduction, increasing speed of recognition does not, by itself, indicate increased automaticity in the process. In this case, we see in Table 2 that the CV data for G.S. also shows a trend to decrease with increasing word frequency. Analysis of the variability reflected in these CV data, by means of the O'Brien (1981) transformation of the data described earlier, indicates that indeed G.S. was showing a significant decrease in variability, over and above any decrease normally associated with a speed-up effect. This supports the conclusion that not only was G.S.'s command of vocabulary strong (as evidenced by the relatively low error rate for both base words and nonwords) but also that automaticity of word recognition was greater for more common words than for less common words. The absence of a sessions effect for variability suggests that G.S. did not gain in automaticity with respect to the mechanical aspects of performing the lexical decision task (skills associated with reading the screen, pressing the correct button, etc.). The frequency effect for RT variability speaks to automatization of word recognition.

It is interesting to compare G.S.'s results from session 1 with those of the three native speakers. As Table 2 shows, the native speakers appear to be considerably faster overall in their response times, especially with the less frequency words. They each showed a significant frequency effect for RT, indicating that the more common the word, the faster the subject responded. What is especially noteworthy is that here the pattern was not accompanied by a significant frequency effect in the analysis of variability. This was not the case for G.S. The absence of variability differences across the vocabulary is consistent with the expectation that for native speakers, who are relatively highly skilled readers of the language, word recognition will be fairly automatic throughout the vocabulary. As noted above, these native speakers of English did show a frequency effect for RT. Taken together, the presence of an effect for RT but not for variability points to speed-up but not increased automaticity for highly frequent words compared to rarer words in these native speakers. The variability effect in the data of the second-language speaker of English, on the other hand, suggests, that for him, there was room to develop word recognition automaticity further.

The second main focus of the study was to investigate the degree to which purposeful reading between testing sessions would affect word recognition skills for selected words. The analyses showed that while RT decreased as a function of session, the variability of RT significantly decreased as a function of session for studied words but not for control words. This result supports the idea that purposeful reading of selected words will change the nature of the underlying recognition process – making it more automatic. This conclusion must, of course, be treated as tentative since the different pattern for studied and control words did not reach statistical significance in the interaction term when the data were analysed together. Future testing of this question should increase the power of the analysis by including many more studied and control words.

In conclusion, this preliminary work has shown the viability of a method of testing for automaticity of word recognition in a single individual with the use of a simple, timed lexical decision task. This work opens up some exciting possibilities for future single-case research in the area of vocabulary development. For example, future research on the role of purposeful reading as a training method could include control tasks in which control words are encountered in reading but where the purpose is not genuinely communicative (e.g., letter-crossing task). In this way one could better test whether purposefulness – or genuinely communicative nature – of the training task is the essential factor that promotes vocabulary recognition skill. Our method of analysing variability in response time data holds promise for a wide range of research on individual differences and development patterns in the acquisition of second-language vocabulary and other complex skills.

IV References

Besner, D. and **Johnston, J.C.** 1989: Reading and the mental lexicon: on the uptake of visual information. In Marslen-Wilson, W., editor, *Lexical representation and process*. Cambridge, MA: MIT Press.

Colley, A. 1987: Text comprehension. In Beech, J.R. and Colley, A.M., editors, *Cognitive approaches to reading*. New York: Wiley.

Cooksey, R. and **Freebody, P.** 1987: Aspects of a computer-managed test of children's reading vocabulary: reliability, validity and characterization of knowledge.*Reading Psychology: An International Quarterly* 8, 103–18.

Craik, F.I.M. and **Lockhart, R.** 1972: Levels of processing: a framework for memory research. *Journal of Verbal Behaviour* 11, 671–84.

den Heyer, K., Goring, A., Gorgichuk, S., Richards, L. and **Landry, M.** 1988: Are lexical decisions a good measure of lexical access? Repetition blocking suggests the affirmative. *Canadian Journal of Psychology* 42, 274–96.

Favreau, M., Komoda, M. and **Segalowitz, N.** 1980: Second language reading: implications of the word superiority effect in skilled bilinguals. *Canadian Journal of Psychology* 34, 370–80.

Favreau, M. and **Segalowitz, N.** 1983: Automatic and controlled processes in the first and second language reading of fluent bilinguals. *Memory and Cognition* 11, 565–74.

Fodor, J.A. 1983: *Modularity of mind.* Cambridge, MA: MIT Press.

Forster, K. 1989: Basic issues in lexical processing. In Marslen–Wilson, W., editor, *Lexical representation and process.* Cambridge, MA: MIT Press.

Gatbonton, E. and **Segalowitz, N.** 1988: Creative automatization: principles for promoting fluency within a communicative framework. *TESOL Quarterly* 88, 473–92.

Graesser, A.C., Hoffman, N.L. and **Clark, L.F.** 1980: Structural components of reading time. *Journal of Verbal Learning and Verbal Behaviour* 19, 135–51.

Graves, M., Ryder, R., Slater, W. and **Calfee, R.** 1987: The relationship between word frequency and reading vocabulary using six metrics of frequency. *Journal of Educational Research* 81, 81–90.

Hashtroudi, S. 1983: Types of semantic elaboration and recall. *Memory and Cognition* 11, 476–84.

Henderson, L. 1982: *Orthography and word recognition in reading.* New York: Academic Press.

Herrington, R., Kennelly, K., Hall, C. and **Baggett, M.** 1994: Practice and the lognormal distribution. Presentation at the TENNET V conference, Montreal, Quebec.

Howell, D.C. 1992: *Statistical methods for psychology* (3rd edn). Boston, MA: PWS-KENT.

Jacoby, L. 1991: A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language* 30, 513–41.

Levene, H. 1960: Robust tests for the equality of variance. In Olkin, I., editor, *Contributions to probability and statistics.* Palo Alto, CA: Stanford University Press, 278–92.

Meara, P. and **Buxton, B.** 1987: An alternative to multiple choice vocabulary tests. *Language Testing* 4, 142–54.

O'Brien, R. 1981: A simple test for variance effects in experimental designs. *Psychological Bulletin* 89, 570–74.

Perfetti, C. 1985: Reading ability. New York: Oxford University Press.

Perfetti, C. and **McCutchen, D.** 1982: Speech processes in reading. In Lass, N., editor, *Speech and language. Vol 7.* New York: Academic Press, 237–69.

Quinlan, P.T. 1992: *The Oxford psycholinguistic database.* Oxford: Oxford University Press.

Segalowitz, N. 1986: Second language reading. In Vaid, J., editor, *Language processing in bilinguals: psycholinguistic and neuropsychological perspectives.* Hillsdale, NJ: Lawrence Erlbaum Associates, 3–19.

—— 1991: Does advanced skill in a second language reduce automaticity in the first language? *Language Learning* 41, 59–83.

Segalowitz, N. and **Hébert, M.** 1990: Phonological recoding in the first and second language reading of skilled bilinguals. *Language Learning* 40, 503–38.

Segalowitz, N., Poulsen, C. and **Komoda, M.** 1991: Lower level components of reading skill in higher level bilinguals: implications for reading instruction. *AILA Review* 8, 15–30.

Segalowitz, N.S. and **Segalowitz, S.J.** 1993: Skilled performance, practice, and the differentiation of speed-up from automatization effects: evidence from second language word recognition. *Applied Psycholinguistics* 14, 369–85.

Stanovich, K. 1980: Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly* 16, 32–71.

—— 1991: Word recognition: changing perspectives. In Barr, R., Kamil. M.L., Mosenthal, P. and Person, P.D., editors, *Handbook of reading research. Volume II.* New York: Longman, 418–52.

Taft, M. 1991: *Reading and the mental lexicon.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Tulving, E. 1979: Relation between encoding specificity and levels of processing. In Cermak, L. and Craik, F., editors, *Levels of processing in human memory.* Hillsdale, NJ: Lawrence Erlbaum Associates, 405–28.

Watson, V. 1993: The development of skill: the coefficient of variation as an index of skill acquisition. Unpublished MA thesis, Concordia University.