Kempe, V., & MacWhinney, B. (1996). The crosslinguistic assessment of foreign language vocabulary learning. *Applied Psycholinguistics, 17*, 149-183.

Abstract

This study explores the use of task that can be applied in a uniform fashion across different languages to compare levels of vocabulary development in foreign language learning. Experiment 1 tested native speakers of Russian and German and demonstrated the basic comparability of subjects' judgments for both words and nonwords. The results for Russian showed an influence of word length which can be understood in terms of the Orthographical Depth Hypothesis. Experiment 2 applied the same task to learners of Russian and German and found that learners of Russian had achieved a lower level of vocabulary control than learners of German at comparable language exposure levels. This disadvantage for Russian can be attributed to the novelty of the Cyrillic graphemic system which restricts the accessibility of written language input at early stages. There was a nonlinear increase over time in word sensitivity that can be attributed to the increasing contribution of lexical plausibility factors at later stages of learning. Also, the lexical decision task appeared to be sensitive to inhibitory effects of concurrently studied languages, as well as decay due to lack of regular exposure. Together, these results indicate that the lexical decision task can be a useful tool for the assessment and crosslinguistic comparison of lexical development in foreign language learning.

Language teachers have long understood that some languages are more difficult than others. To properly assign students to languages at both universities and large language teaching institutes, it is important to understand the exact nature of differing challenges that languages present to the student (MacWhinney, in press). In order to better understand these issues, researchers interested in foreign language acquisition (FLA) need to perform crosslinguistic comparisons. However, these comparisons can only make sense if we can compare groups that are at equal levels of language exposure. For example, we might want to know whether second-year learners of German have learned more German than second-year learners of Russian have learned Russian, given comparable levels of exposure? More specifically, we might want to know whether second-year learners of German have acquired a larger vocabulary or have more accurate pronunciation than second-year learners of Russian, given comparable levels of exposure.

If we want to conduct crosslinguistic experimental examinations of FLA, we need to find some way to equate level of attainment across learners of different languages. And we want to have ways of assessing proficiency that can be meaningfully interpreted in psycholinguistic terms. This paper explores a new methodology for developing such comparisons in the area of lexical learning and applies these techniques to a comparison of lexical learning in German and Russian. There are four reasons why current methodology needs to be adapted to provide a more solid basis for crosslinguistic studies.

1. The first problem lies with the common use of the **duration** of exposure as a surrogate variable for **amount** of exposure. Virtually all studies that make an attempt to assess different levels of learner ability share this weakness, whether they are conducted in the context of the Competition Model (Kilborn & Ito, 1989; McDonald, 1987), criticial period theory (Flege & Davidian, 1984; Johnson & Newport, 1989; Johnson & Newport, 1991), or Universal Grammar (White, 1990; White, 1991). The use of duration of exposure as a surrogate for amount of exposure is never defended on principled grounds. Virtually any standard theoretical perspective would assume that amount of exposure is more important than duration of exposure as a predictor of the degree of FL learning. However, there is good reason to believe that duration of exposure does not correlate tightly with the actual total amount of exposure. The amount of actual "time on task" is determined by the nature of the individual learner, the structure of the language program, the frequency of opportunities for interaction, and a host of other variables that are typically never measured or even examined. Thus, duration of exposure cannot reliably estimate amount of exposure.
2. Even if we had clear estimates of the amount of exposure to the FL, it would be dangerous to assume that similar amounts of exposure would lead to similar learning outcomes. Differences in learner backgrounds, the shape of target structures in the FL, possible L1 to FL transfer effects, and other factors can all lead

to important differences in the level of FL attainment that are not closely related to the amount of exposure to the FL. For example, learners of Russian must confront a new system of orthographic characters before they can even begin to read new words in the language. This will, in turn, delay the effective acquisition of new lexical items through textbooks and other printed sources. No such problem presents itself for learners of languages that use the Roman alphabet.

3. A third problem involved in providing a more solid basis for crosslinguistic comparisons involves access to crosslinguistically comparable proficiency tests. In principle, we would like to use proficiency test scores as covariates in crosslinguistic comparisons. However, solid proficiency tests only exist for high-profile languages like English, Spanish, German, and French. It is extremely difficult or perhaps impossible to find solid, recognized proficiency tests for even such important languages as Danish, Hungarian, or Vietnamese. Even when they are available, there is almost never a comparable form of the test for two different languages. For example, there are no comparable proficiency tests for German and Russian, or even for English and Russian. Paradis and Libben (1987) have developed an instrument for the study of relative levels of impairment in bilingual aphasics called the BAT (Bilingual Aphasia Test). However, this test can only be used to make comparative assessments of degree of impairment within individual subjects. It is not designed for use in group studies and cannot be used for crosslinguistic experimental work. In addition, it would be dangerous to attempt to make any direct application of a test developed for language-disordered bilinguals to groups of non-disordered non-bilinguals.

4. A fourth problem involves the mismatch between the measures provided by language proficiency measures and those needed for the prediction of outcomes in experimental studies. Proficiency tests typically include the assessment of a wide range of abilities including reading, writing, speaking and listening, as well as the

> assessment of various language domains such as vocabulary knowledge, syntax, morphology and phonology. However, experimental studies of FLA usually focus on very specific aspects of language performance such as grammaticality judgments or reaction times in sentence comprehension tasks. If proficiency tests yielded a full set of subtest scores, we could use these individual components as predictors of outcomes in experimental tasks. However, often subtest scores are not available, or are not clearly associated with specific components of language learning.

For crosslinguistic comparisons, we need measures which estimate the degree of the learner's familiarity with specific FL skills without depending on measures of duration of exposure. These measures should be restricted to one specific language domain and should be easy to construct and easy to apply in a uniform fashion across languages.

In this paper, we choose to focus our attention on the evaluation of lexical learning. We understand that a full crosslinguistic assessment of differences in target language difficulty will require the construction of multiple assessment insturments in each of the many subdomains of language performance(MacWhinney, in press). However, we must begin work toward this larger goal by delineating a specific, psycholinguistically manageable area for assessment. We have chosen the lexical domain for two reasons. Firstly, we believe that lexical learning is central to many aspects of early second language learning (MacWhinney, 1992). Lexical items provide an initial focus for learning in phonology, orthography, syntax, and semantics. Secondly, we believe that good experimental tasks are available for the crosslinguistic measurement of levels of lexical learning. One such task is the lexical decision task, which we have utilized in the current study.

The Lexical Decision Task

The task that we have chosen as a basis for our study is the standard lexical decision task that has been used so widely in psycholinguistics (Forster, 1976; Forster, 1985). In this task, learners are presented with a large list of FL words and nonwords and have to decide for each item whether or not it is a word that they know. If they know the word they should say “yes”. If they did not know the word, they should say “no.” Meara and Buxton (Meara & Buxton, 1987) report an first version of this task, which served as the initial impetus to our work.

One advantage of the lexical decision task over standard multiple-choice vocabulary tests is that the number of words presented in one session can be increased substantially, thus making the estimations much more reliable (Anderson & Freebody, 1983). The second advantage of a test based on the lexical decision task is that it minimizes the effects of strategies and contextual influences on test taking behavior. Anderson and Freebody (1983) carefully compared performance on a lexical decision vocabulary test and a multiple choice test, both containing the same test words, in 120 native 5th-graders. Although there was a strong correlation between the two tests, the authors showed that subjects were more likely to really know the meaning of words which were indicated as familiar in the lexical decision test than of words for which they choose the correct alternative in the multiple choice test. This is because multiple-choice behavior is more affected by extraneous factors, such as test-taking strategies, not related to vocabulary knowledge.

In order to decide whether a given input pattern is a word, the subject can evaluate a wide variety of orthographic, phonological, semantic, and contextual cues. There is evidence that subjects can vary the relative weights of these cues according to the actual task demands (Balota & Chumbley, 1984; Gordon, 1983; Seidenberg & McClelland, 1989; Stone & Van Orden, 1993). For example, the two-stage-model proposed by Balota and Chumbley assumes that stimuli vary in terms of their strength on a familiarity/meaningfulness continuum and that subjects establish an upper bound for a “yes” decision and a lower bound for a “no” decision. When familiarity/meaningfulness values fall below the lower bound, “no” responses can be given quickly; when these values fall above the upper bound, “yes” decisions are fast. However, when familiarity/meaningfulness values fall in the region between the two bounds, additional cues must be considered, and reaction times will increase. Word frequency effects can be attributed to the fact that the proportion of low frequency words falling in the region of further analysis between the two criteria is higher than the proportion of high frequency words. Consequently, variations of the nonwords alone can modulate subjects' sensitivity to word-nonword distinctions and, thereby, the magnitude of the word frequency effect. The more distinct the words and nonwords in the experiment, the better subjects can discriminate between them and the fewer will be the low frequency words that fall into the region of further, slower analysis.

The Balota and Chumbley model presumes a stage of explicit lexical access and a postlexical decision stage. An alternative view is expressed in the interactive activation model of Seidenberg and McClelland (1989) that permits all types of information to contribute interactively during the progressive activation of lexical forms. Rather than focusing on a particular moment of lexical activation, this model views activation as growing over time. However, both models recognize the importance of a threshold level for final production of a "yes" or "no" response.

In the interactive activation framework, performance on the lexical decision task can be facilitated by information about a word's orthography, morphological form, meaning, or sound. Cues deriving from each of these information sources can be assumed to develop jointly since they all are linked to language experience and determined by the processed language input. However, sometimes it is possible to give evidence for the importance of one of these data sources over others. For example, Anderson & Freebody (1983) showed that among fifth graders, the subjects with the highest proportion of correctly identified words had also the highest rate of false alarms on pseudoderivatives, i.e. nonwords like "observement" which combine real English morphemes in illegal combinations. The false alarms to these pseudoderivatives reflects the students' occasionally inappropriate use of their growing knowledge of English morphology. In these cases of false alarms to pseudoderivatives, the subject is responding not to the semantic knowledge of the word, but to its plausibility as an English word. On the other hand, subjects can easily reject words like “svliz” on the basic of their violation of recognized phonological patterns. Whether the two sources, semantic knowledge and lexical plausibility, are in coalition or competition with each other depends on the stimulus material.

It can be assumed that, as learning progresses, word and nonword distributions generally shift apart. However, the increase in word sensitivity should be restrained to the extent that lexical plausibility comes into play. We will define “lexical plausibility” as the computation of evidence that a possible lexical item matches the morphological, orthographic, and phonological patterns implicit in the lexicon. Thus, we would expect the increase in sensitivity to be better described by a non-linear rather than a linear relationship where the acceleration slows down with increases in the match between nonwords and the phonological, orthographic, and morphological characteristics of words implicit in the lexicon. This hypothesized pattern is illustrated in Figure 1.

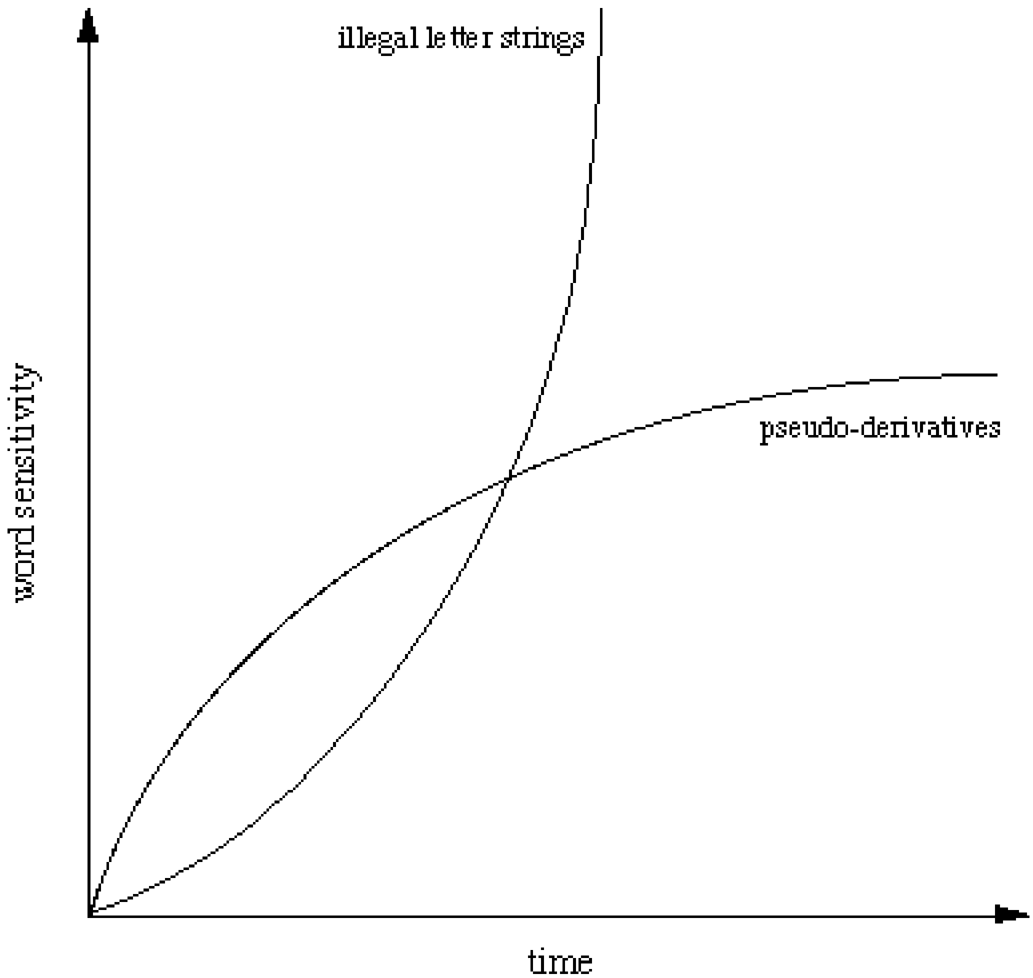

Figure 1: Hypothetical development of word sensitivity over exposure time in FL-learners if the stimulus material contains (a) letter strings illegal in a given FL, or (b) orthographically and morphologically close nonwords like pseudo-derivatives.

The two curves in Figure 1 illustrate hypothetical relations between the increase of word-sensitivity and exposure time for different stimulus materials. Salient differences between words and nonwords, like the presentation of illegal letter strings, will overestimate the learner's lexical knowledge; whereas subtle differences, like the presentation of pseudoderivatives, will underestimate it. This, however, does not preclude the possibility of crosslinguistic comparisons as long as the word and nonword distributions are equidistant across languages. This equidistance can be achieved by using identical sampling and matching criteria in each language. Our attempt to find and test such a sampling and matching algorithm is the basis of Experiment 1.

Experiment 1 - The native speaker baseline

The two languages we selected for this crosslinguistic comparison of FL learning were German and Russian. In order to compare lexical learning in these two languages, we needed to construct two distributions of test items in which the distance of the words from the nonwords and the relative familiarity of the two sets of materials were similar across the two languages. In this first study, we constructed these two distributions and tested their equivalence by having native speakers perform the lexical decisions task. This not only provides a native speaker baseline of performance against which FL-learners can be compared, but also allows us to evaluate the equivalence between the word and nonword distribution across languages. The evaluation of equivalence is based on three measures:

1. the d-prime sensitivity measure,
2. the decision times for words and nonwords, where equal differences between word and nonword decision times in both languages again would suggest equivalence of the distributions and,
3. the magnitude of frequency effects on decision times for words in both languages which provide information about the comparability of the normative frequency counts.

If these three measures do not differ substantially, we can consider the test materials to be roughly equivalent in the two languages with respect to lexical processing by native speakers.

Method

Word Stimuli

In order to guarantee equivalence between the word and nonword distributions in these two languages, we choose words on the basis of a complete frequency count for each of the two languages. Normative frequencies were obtained from the “Frequency Dictionary of Russian” (Zasorina, 1977) and from the CELEX-Database (Baayen, Piepenbrock & van Rijn, 1993) for German. Both counts are based on samples derived from a variety of types of written text. The Russian count is based on a corpus of approximately one million words and the German count is based on a corpus of nearly ten times that amount. We computed the logarithm of the lowest and highest frequency forms in each corpus, which resulted in log frequencies between -.0012 and 4.5. This frequency continuum was then divided into 240 equal segments resulting in a log-unit of 0.0188. In other words, the whole frequency continuum was linearized and segmented into 240 discrete points. For each point, the expected frequency per million was computed and assigned its closest frequency matching word in both languages. Between languages, each pair of words at a certain expected frequency value was additionally matched for word class as close as possible. For example, if a noun with a frequency of 1.99 tokens per million was selected in German and no noun but only a verb with this frequency was available in Russian, then the verb was selected instead but at the next point the German target word then would be a verb and the Russian counterpart a noun. This led to very similar distributions of word classes over word frequencies for both languages. Above a frequency of 1448 tokens per million (after 181 log-units) it is not possible to find matching words for all subsequent expected frequencies. Thus, we stopped the selection at this point. The outcome of the described sampling procedure is a pool of 181 words in each language with steadily increasing log-frequencies which are also matched for frequency and word class between the two languages.

Normative word frequencies here serve only as an approximation of frequencies of occurrence in the input of FL-learners. Since one can assume some differences between the language input encountered by native speakers and the one encountered by FL-learners, it is necessary to perform an additional check of whether normative word frequencies are indeed a good approximation of FL-input frequencies.

Comparison of L1 and FL word frequencies

Comparability of the stimulus material presupposes that the likelihood of occurrence in the input to FL learners is similar between the two languages. But how well does FL input indeed reflect native language frequency distributions, particularly, if people acquire the language in a formal setting and through formal tools as language textbooks? While some textbooks rely on frequency counts in determining which vocabulary items to include in the text (Keller, 1991), others do not. Although many European language programs have established vocabulary minima that could be used to indicate the shape of the input to the FL learner, no such accepted minima exist in the United States. The lack of established minima and the absence of standards for textbook construction makes it difficult to assess exactly what vocabulary a learner in a American classroom might have encountered. If we are interested in conducting research on FL learning in the United States, we will need to find a way of assessing the lexical distribution characteristics of the input.

To obtain a rough estimate of whether the frequencies of the selected word pool correlate with the frequencies of occurrence in language textbooks, we consulted the vocabulary lists of five widely used language textbooks for each language. The references for the textbooks are given in Appendix 2. For each word, we counted the number of occurrences across the five lists. These numbers ranged from 0 to 5. The correlations between real-life word frequency and the frequency of occurrence in language textbooks are .56 ($p < .001$) for German and .68 ($p < .001$) for Russian. This indicates that textbook vocabulary reflects native word frequency quite well in that the more frequent a given word the more likely it is to occur in a textbook and, consequently, in the input to the learner.

Nonword Stimuli

In order not to overestimate the learner's lexical knowledge we decided to maintain a certain degree of similarity between words and nonwords. This was achieved by using the following criteria:

1. Pseudohomophones as well as cognates of English words were excluded.
2. The nonwords had to obey the phonotactic rules of the language.
3. Each nonword had to match the overall phonological shape, word class, and derivational type of its word counterpart. For polymorphemic words, the matching nonword had to have a phonological shape that was similar in terms of syllable structure, overall length, and phonemic complexity.
4. If the word was a derivative, the matching nonword was constructed to look like a plausible derivative, or what we call a pseudoderivative. For example, the German derivation “Angriff” *(assault)* was paired with the pseudoderivative “Aufplatz” and the compound “Wettlauf” *(race)* was paired with the pseudoderivative “Topfgang”. The forms “Aufplatz” and “Topfgang” consist of really existing suffixes and stems yet these particular combinations do not exist in German. These nonwords are henceforth called pseudoderivatives. For Russian, analogous examples are given by the word “krasnorecie” *(eloquence)* and the pseudoderivative “uzkocuvstvie”. The inclusion of pseudoderivatives is essential for two reasons: First, without them, subjects might discover that all items containing derivational morphemes are actually words which would considerably overestimate their lexical knowledge. Second, pseudoderivatives permit a very close match of orthographical and morphological similarity between targets and distractors, thereby improving both the reliability and difficulty of the test.

This matching procedure resulted in 52 pseudoderivatives for German and 70 for Russian. Given the large size of the word pool, the different proportions of pseudoderivatives can be taken to accurately reflect the different amount of derivational morphology in German and

Russian. The complete set of words and nonwords for each language is given in Appendix 1.

In each language, 181 words were selected according to the procedure described above. For each word, a nonword match was constructed following the criteria described above. Thus, the stimulus material consisted of a total of 362 items in each language. However, due to mistypings in the actual presentation, the analysis excluded one word in the Russian version and one nonword in the German version

Subjects

There were 28 subjects in this experiment. They were recruited by advertisement and paid $5.00 for their participation. There were 13 native speakers of German studying or working in Pittsburgh (10 men, 3 women). There were 15 native speakers of Russian (8 men, 7 women). Seven of the Russian subjects were immigrants and the remaining 8 were participants in a student exchange program. All but one subject in the Russian group had received partial or complete college or university education in their native language.

Since all subjects were living in an English environment at the time of testing we wanted to make sure that none of them had become English dominant. To determine this, all subjects completed a questionnaire prior to the experiment which included questions about the length of their stay in the United States, length of formal instruction in English, age of first exposure to English, an estimation of their average use of different languages per day at the present time, and, finally, knowledge of other non-native languages. Also, subjects were asked to provide self-ratings of their speaking, reading, writing and comprehension abilities in English on a scale from 1 to 6. These data are summarized in Table 1.

As can be seen from Table 1, both groups were homogeneous with respect to age and stay in the United States, but differed in the extent to which they knew and used English, with the Germans making more day-to-day use of English than the Russians.

	German (n = 13)	Russian (n = 15)
age in years	26.2 (3.5)	26.3 (6.4)
length of stay in the US in months	9.0 (6.7)	13.3 (12.1)
estimated % of L1-usage per day**	27.5 (27.9)	75.4 (23.5)
years of English instruction**	9.1 (2.3)	4.4 (4.8)
average self-rating for English*	4.3 (.7)	3.3 (1.1)
age of first exposure to English	10.8 (1.8)	13.8 (8.9)
other non-native languages studied or spoken longer than 1 year	1.2 (.6)	.6 (1.1)

** both groups differ significantly (p <.01)
* both groups differ significantly (p <.05)

Table 1: Summarized questionnaire data for the two groups of native speakers. All cells include means with standard deviations in parentheses.

Procedure

Subjects first completed the questionnaire. All questionnaires, instructions, and stimuli were presented in the subject’s native language. After the questionnaire was complete, subjects were seated in front of a Macintosh Centris 660AV. Items were presented in black characters on a light blue screen covering a visual angle of approximately 2.5 degrees. Subjects were instructed to read the items and to decide whether they know the word or not. This instruction was preferred to the typical word-nonword instruction in lexical decision in order to keep the conditions for native speakers and FL-learners equal. Subjects responded by pressing the appropriate buttons on the CMU button box.

Each item was preceded by a 200 msec. ISI and was visible on the screen until the subject made the decision. The dependent variables were the correctness of the decision and the reaction time measured from the onset of the display of the item on the CRT. Each subject saw the whole set of stimuli in an individually randomized order. A session lasted for approximately 10 to 15 minutes.

Results

Error data

For each subject, the proportions of hits and false alarms were used to calculate a d-prime score which was then entered in a one-way ANOVA with Language as a between-subject variable. This analysis revealed no significant effect of Language ($p > .1$). The group means for the d-prime values are 3.720 for German and 3.455 for Russian.

To evaluate whether the effect of orthographical and phonological similarity was identical in both languages, we need to examine the proportion of the false alarms that are due specifically to pseudoderivatives. However, this proportion must be corrected for differences in the base rates of occurrence of pseudoderivatives in the total stimulus set in the two languages. There were 70 pseudoderivatives in the Russian stimulus set out of a total of 181 nonwords and 52 in the German set. These numbers represent percentages of 39% and 29% pseudoderivatives for Russian and German, respectively. If subjects were to perform entirely at a chance level, we would expect that the Russian speakers would have 39% of their false alarms for pseudoderivatives and that the German speakers would have 29%. We subtracted these chance levels from the actual proportions in order to obtain a corrected false alarm rate for pseudoderivatives. A planned comparison showed that the Russian group had a significantly lower proportion of errors on pseudoderivatives (.29 above chance) than the German group (.53 above chance) (Scheffe's test, $p <.01$). Thus, pseudoderivatives induced error rates above chance level in both languages, but the errors were more concentrated on pseudoderivatives for Germans than for Russians, although the false alarm rates across all items were similar.

Reaction times - ANOVA analyses

Values beyond two standard deviations from a subject's mean were truncated. This procedure affected 3.5% and 4.2% of the data points for Germans and Russians, respectively.

We conducted a 2 (Language) x 2 (Lexicality) x 2 (Derivation) ANOVA on the reaction time data. The cell means are shown in Table 2. Note that the reliability of the subject means for derivatives and non-derivatives is different since their proportions were slightly different between languages. This analysis revealed significant main effects of Lexicality ($F_s(1,26) = 129.1$, $p < .001$; $F_i(1,714) = 913.18$, $p < .001$) and of Derivation ($F_s(1,26) = 233.1$, $p < .001$; $F_i(1,714) = 328.74$, $p < .001$), suggesting that reaction times are slower for nonwords and for derivatives. The two-way interaction of Lexicality and Derivation ($F_s(1,26) = 94.5$, $p < .001$; $F_i(1,714) = 12.17$, $p < .001$) indicates that the impact of Derivation on reaction times is much higher in nonwords. The significant three-way interaction of Lexicality, Derivation and Language ($F_s(1,26) = 7.4$, $p < .05$; $F_i(1,714) = 8.24$, $p < .01$) shows that this pattern of results is somewhat different in the two languages.

Planned comparisons between languages revealed an difference that failed to reach significance for non-derivative nonwords (German: 889 ms, Russian: 1029 ms) ($t(26) = 1.79$, $p < .1$) where reaction times are slower in Russian. All other differences were nonsignificant (all p's > .4). This demonstrates that the processing of simple nonwords, and not pseudoderivatives, is slightly slower in Russian. One possible reason for this slow-down in simple nonwords in Russian could be their length. Our frequency-based procedure for construction of the two lists did not attempt to balance for length and there is a general tendency for non-derived Russian words to be a bit longer than German words. These differences could be the cause of the reaction time asymmetries.

		German	Russian
words	non-derivatives	764	739
	derivatives	805	843
	all words	772	775
nonwords	non-derivatives	889	1026
	derivatives	1201	1282
	all nonwords	959	1119

Table 2: Mean reaction times in milliseconds for non-derivatives and derivatives in the Russian and German versions of the lexical decision task (native speakers). The figures for "all words" are not the average of the separate cell means, since there were more non-derivatives than derivatives.

Reaction times -- ANCOVA analyses

To examine possible word length effects, we included Length as a covariate in a three factors (Language, Lexicality, and Derivation) ANCOVA by items. This analysis yielded a marginally significant main effect of Language ($F(1,706) = 3.72$, $p = .05$) as well as a significant effect of the covariate of Length ($F(1,706) = 95.0$, $p < .001$). The tests for common slopes of the regression lines on different levels of the three factors were significant for the 2-way interactions of Length by Language ($F(1,706) = 10.1$, $p < .01$), Length by Lexicality ($F(1,706) = 14.4$, $p < .001$), Length by Derivation ($F(1,706) = 3.8$, $p = .05$), as well as for the 3-way interaction of Length, Lexicality and Derivation ($F(1,706) = 7.2$, $p < .01$). No other effects reached significance. This model accounted for 73.3% of the variance in the reaction times. Figure 2 illustrates the differences in the slopes of the effects of word length on the various factor levels.

As becomes apparent from this analysis, differences in the mean reaction times for different factor levels are entirely due to differences in the impact of word length. First, as the Length by Language interaction shows, reaction times in both languages are to a considerable degree influenced by word length, but this influence is much larger in Russian. This is indicated by the fact that the slope of the reaction times over word length is generally higher in Russian. Second, the slope differs also with respect to stimulus type with higher slopes in the nonwords. Third, the Length by Derivation interaction reveals an increase in the slope for derivatives which is further modulated by the three-way interaction of Length, Stimulus and Derivation. This three-way interaction is due to the fact that the higher increase of reaction times with word length is mainly due to pseudoderivatives. The main effect of language, which is a test of common intercepts, is a consequence of the overall higher slope of word length in Russian and indicates only that the intercept for reaction times in Russian is below and not above the one in German. We therefore conclude that the larger difference in the mean reaction times between words and nonwords in Russian is due to a word length effect.

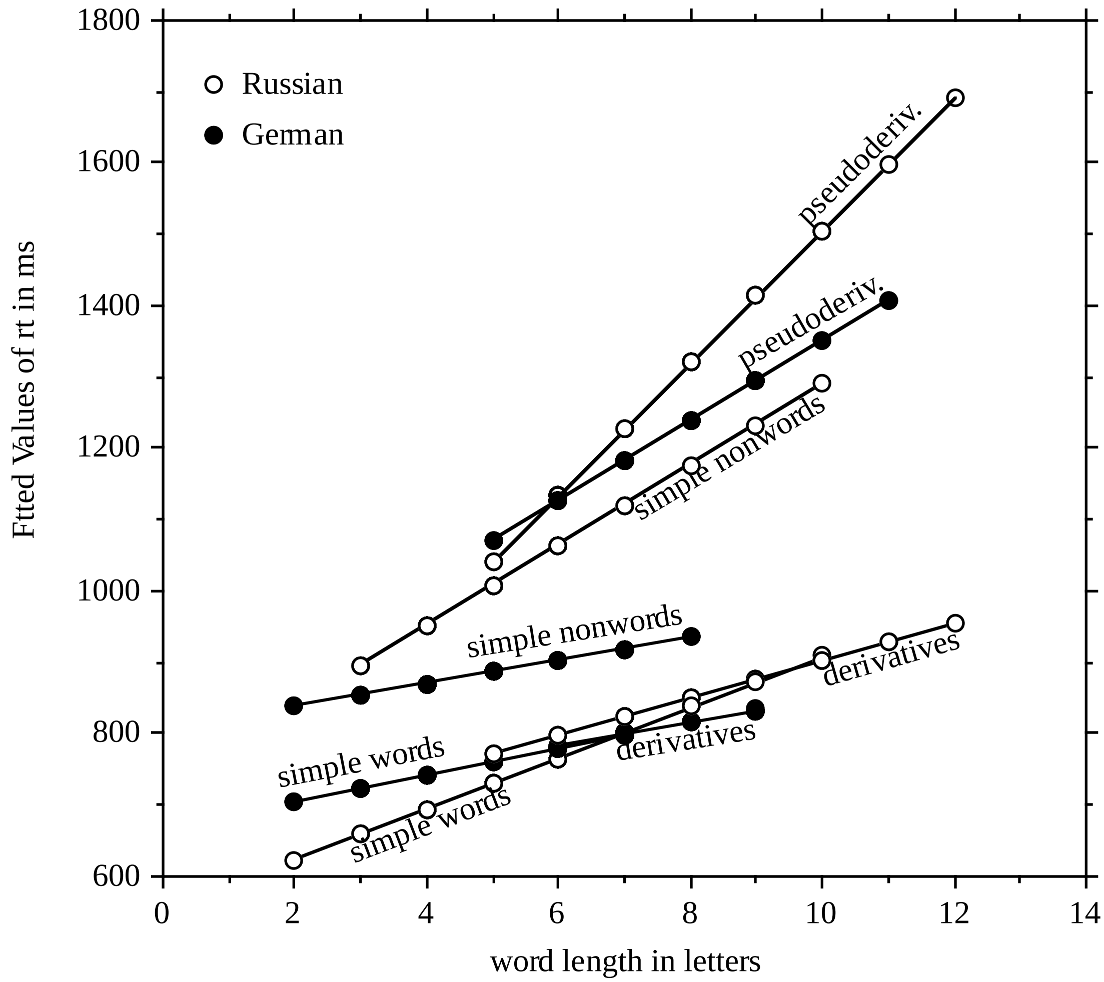

Figure 2: Fitted values of reaction times over word length for derived and underived words and nonwords in Russian and German native speakers.

Reaction times -- HRA analyses

In order to obtain a confirmation for the compatibility of the frequency-based word sampling procedure between Russian and German, we wanted to determine whether there are any language-dependent differences in the effect of word frequency on the reaction times. To measure this, we conducted a hierarchical regression analysis on the reaction times for the words in each language to determine the portion of variance accounted for by word frequency over and above the effect of word length. Word length was entered at the first step in the equation, followed by logWF (logarithm of word frequency per million). The interaction of both was entered at the last step.

As can be seen from Table 3, logWF accounted for a significant proportion of variance in both languages (12% for Russian learners, 5.6% for German learners). The interaction of word length and logWF was not significant in either language, although logWF accounted for twice as much variance in Russian as in German.

To check whether this difference is reliable we performed another regression analysis on the words of both languages with language as an additional independent variable. Length, logWF, language, and the interactions of Length with logWF and Length with Language were entered as a block at the first step. The interaction of logWF and Language was entered at the second step to determine whether this interaction accounts for a significant proportion of variance over and above all other independent variables and their interactions. The Frequency x Language interaction accounted for an additional 0.8 % of variance which was marginally significant ($F(1,345) = 3.63$, $p = .06$). The sign of the regression coefficient indicates that the word frequency effect tended to be slightly stronger in the Russian group.

Language	Variable	beta weights	cumulative r^2	F change
Russian				
	WL	35.73	.287	71.61 **
	WFreq	-46.51	.407	37.31 **
	WL x WFREQ	-6.85	.417	3.01
German				
	WL	18.25	.088	17.34 **
	WFreq	-25.34	.144	11.58 **
	WL x WFREQ	-4.93	.150	1.24

** $p < .01$

Table 3: Stepwise regression of word length and word frequency on the reaction times for Russian and German words. The order of independent variables in the table correponds to the order of entry in the regression.

Discussion

The purpose of this experiment was to establish a native speaker baseline for both languages and to examine whether the selected word material yielded similar results with respect to error rates and the size of the lexicality and word frequency effects.

The absence of significant language dependent differences in d-prime suggests that our item selection procedure yields a set of test items that are comparable between languages with respect to lexical decision difficulty. In other words, native speakers of both languages are equally sensitive to the word-nonword distinction as employed in our test material. The language differences for word frequency effects on reaction times were not significant, but there was a slight tendency for a stronger word frequency effect in Russian. This lack of significant differences supports the general success of the attempt to prepare roughly equivalent versions of the measure in the two languages at least in terms of the use of this measure in studies of foreign language learning.

Error rates on nonwords suggested that the Russian simple nonwords might have been slightly less discriminable from words than the German ones. However, the similar magnitude of d-prime and the reaction times suggested instead that differences in performance on simple nonwords were a result of differences in word length. After partialling out the effect of word length and its interaction with language, no differences in the lexicality effect remained. This indicates that the distributions of words and nonwords in both languages are by and large equidistant. We therefore would like to conclude that our matching procedure was fairly compatible between Russian and German and that the imbalance in the distribution of false alarms might reflect differences in saliency of derivational morphology in that German pseudoderivatives tend to produce a higher false alarm rate.

The finding that the word length effect was significantly stronger in Russian is worth considering in greater detail. Dual-route theories of reading (Coltheart, Curtis, Atkins & Haller, 1993) attribute word length effects to the use of a phonological assembly process that constructs the word by repeated application of successive grapheme-phoneme correspondences. Several authors (Feldman & Turvey, 1983; Frost, 1994; Frost, Katz & Bentin, 1987; Lukatela, Popadic, Ognjenovic & Turvey, 1980) have argued that, in languages with shallow orthographies, where there is a direct and unequivocal relationship between graphemes and phonemes, readers tend to maximize use of this phonological assembly route. In languages with deep orthographies, where the grapheme-phoneme relations are less transparent, the application of conversion rules is not efficient, and readers are forced to favor direct lexical access.

In terms of consistency of grapheme-phoneme relations, both German and Russian can be considered to be languages with shallow orthography[1]. However, Russian and German differ in the ways in which grapheme-phoneme correspondences are structured. In Russian, a single grapheme often represents a phoneme. Typically, a grapheme-phoneme correspondence is matched by the opposite phoneme-grapheme correspondence. In German, the relations between phonemes and graphemes are more asymmetrical than in Russian. If one reads a word in German, there is not much doubt about how to pronounce it. However, if one hears a word, different spelling versions are often possible (of which, of course, only one is correct). This is because many phonemes can be expressed by different grapheme clusters in written German. For example, [a:] can be spelled either as "a" as in "Wal" *(whale)* ,"ah" as in "Wahl" *(election)* , or "aa" as in "Saal" *(room)*. Although there is often a good mapping from a grapheme or grapheme cluster to a phoneme, it is extremely difficult to map from a phoneme back to a grapheme.[2]

Consider what happens when the subject encounters a string like "Waal" (a nonword). By phonological assembly the form [wa:l] becomes activated in the phonological lexicon. However, this information is insufficient to decide whether the string is a word or a nonword and a subsequent spelling verification is necessary. It is this additional step which makes a default use of phonological assembly in German inefficient. Similar considerations hold true for other languages considered to be shallow but containing a considerable amount of homophones as in Spanish (Sebastián-Galles, 1991) or Dutch. A similar proposal about the role of sound-spelling correspondences in word recognition was recently made by Stone (1994). In terms of these various models, it would appear that the presence of a word length effect in Russian, but not in German, would point to a great use of the assembly route in Russian native speakers.

Experiment 2 - Lexical decision performance in learners

Having obtained evidence for the comparability of our stimulus material between Russian and German for native speakers and having explored the potential sources of reaction time differences between the two languages, we are now ready to examine the use of the lexical decision task as a way of measuring relative levels of achievement in foreign language learners. Our major goal in this work is the construction of a system for making consistent, psycholinguistically meaningful, crosslinguistic evaluations of levels of language learning with particular respect to lexical development. In order to achieve this overall goal, we would like to show that our lexical decision measure satisfies these four criteria:

1. The measure must have a consistent relation to increasing levels of language exposure for learners as well as native speakers.
2. The measure should be sensitive to breaks in language exposure during learning and to the competing effects of the learning of other languages.
3. There should be a good correlation between reaction times and d-prime scores for learners as well as native speakers.

4. Finally, d-prime scores should correlate well with other measures of language proficiency or attainment.

In this study of lexical decision in foreign language learners, we are focusing more on lexical sensitivity, as measured by d-prime, than on the speed of lexical access, as measured by reaction time. We know that language learners are slow and highly strategic in this task. They are often unsure of their knowledge and try to make use of all the information at their disposal. Their long reaction times are much like those found in young children (Walley, 1988) or aphasics (Friederici, Weissenborn & Kail, 1991; Wulfeck et al., 1989). Because analyses of modularity (Fodor, 1983) and context effects (Swinney, 1979) during lexical access depend crucially on the exclusion of strategic processing (Balota & Chumbley, 1985), it makes sense to confine such studies to the examination of normal, highly fluent adults, rather than children or second language learners. However, the study of lexical processing in children, aphasics, and second language learners can be studied as a topic in its own right, as long one is careful not to ignore the great differences between highly skilled adults and these less-skilled learner groups.

Method

Materials

The materials were the same as those used in Experiment 1.

Subjects

There were 44 subjects, including 22 learners of German (13 men, 9 women) and 22 learners of Russian (12 men, 10 women). They were recruited by advertisement and were paid $5.00 for their participation. All subjects were either college students or had finished college recently. The mean age of the Russian learners was 21.9 years and the mean age of the German learners was also 21.9 years.

Procedure

All subjects were tested individually. At the beginning of the interview, each subject was given a Language Background Questionnaire which included questions about duration and intensity of both formal and informal foreign language instruction and knowledge of other foreign languages. In addition, subjects were asked to provide a self-rating of their reading, writing, listening and speaking abilities in all studied languages on a scale from 1 to 6. The German learners were given the Computer Adaptive Placement Test (CAPT) to estimate the extent to which our lexical familiarity measure would correlate with a general language proficiency measure[3]. As mentioned above, no comparable test could be found for Russian.

The lexical decision task was applied in exactly the same way as with the native speakers. The only difference was that instructions were provided to the learners in English. Dependent variables included the test score for the CAPT, as well as error rates and reaction times for the lexical decision task.

Results

The answers on the Language Background Questionnaire were coded in terms of the following variables:

(a) Formal study time (FormST). This was the duration of formal FL instruction in months as reported by the subjects.

(b) Informal Study Time (InformST). This was the duration in months of informal FL experiences such as immersion experiences during stays abroad or any other extensive contact with the FL that the subjects themselves considered as having improved their FL proficiency. If subjects acquired the FL in a language course in a FL environment the time was coded as both FormST and InformST.

(c) Total study time (TotalST). This was the total time in months of contact with the FL in both formal and informal settings, which was coded as the sum of FormST and

InformST. Periods of overlapping FormST and InformST were only counted once in computing TotalST.

(d) Breaks. These were coded as the total duration in months of interruptions in the course of language learning.

(e) Recency. This was coded as the time in months elapsed since the subject's last regular exposure to the language. The Recency and Breaks variables were included in order to account for any possible decay of language familiarity due to a lack of regular exposure to the language before the experiment.

(f) Self-rating. This was the mean score across the four areas of reading, writing, speaking, and listening for which subjects provided estimates of their own proficiency.

Also, the total study time for additional FLs as well as the corresponding self-ratings were coded for French, Spanish, and Latin, since these were the most frequent additional FLs in our sample. The remaining languages were summarized into one single variable which will be called "Other FLs".

<u>D-prime and overall exposure</u>

Table 4 presents the means, standard deviations, and the correlation matrix for the exposure time variables, d-prime, self-rating, and reaction times for both groups, as well as the CAPT score for the learners of German.

We conducted two separate polynomial regression analyses for each language. The analysis for German combined native speakers and learners into a single group, as did the analysis for Russian. These analyses used d-prime as the outcome variable and TotalST as the only predictor variable. The model included both a linear and a quadratic term. Exposure time for the native speakers was simply estimated by their age in months. In these analyses, the quadratic term yielded a small but significant effect in each language (German: $F(1,32) = 6.3$; $p < .05$; Russian: $F(1,34) = 8.7$; $p < .01$) suggesting that the increase in lexical knowledge eventually slows down at the highest levels of exposure. The fitted curves for both languages are given in Figure 3.

		d'	Self-Rating	Overall RT (ms)	Form ST	Inform ST	Total ST	Breaks	Re-cency	CAPT
mean	G	1.59	3.12	1393	23.5	3.5	26.5	8.0	13.4	31.0
sd		(.33)	(.93)	(421)	(17.5)	(6.3)	(18.0)	(10.8)	(23.1)	(8.7)
	R	*		* *						
mean		1.29	2.69	1836	22.0	3.4	25.2	7.4	7.4	---
sd		(.51)	(1.05)	(612)	(13.1)	(6.6)	(16.3)	(9.2)	(14.2)	
d'										
Self-Rating	G	.39								
	R	.61**								
Overall RT	G	.03	-.12							
	R	.25	-.42*							
Form ST	G	.33	.37	-.14						
	R	.50*	.61**	-.19						
Inform ST	G	.17	.10	.04	-.13					
	R	.35	.19	-.04	.37					
Total ST	G	.39	.37	-.14	.96**	.14				
	R	.54*	.57**	-.17	.94**	.67**				
Breaks	G	.10	.29	.28	-.27	-.13	-.27			
	R	.20	.17	-.13	.49*	.75**	.67**			
Recency	G	.31	-.18	.32	.02	.28	.02	-.08		
	R	-.01	-.33	.67**	.00	.07	-.02	-.21		
CAPT	G	.60**	.49*	-.27	.26	.35	.26	.21	.31	
	R	--	--	--	--	--	--	--	--	

** $p < .01$; * $p < .05$

Table 4: Means, standard deviations, and correlation coefficients for various measures in the learner groups. For each variable designated in the left column the upper row (marked with G) represents the learners of German and the lower row (marked with R) the learners of Russian. All time variables are given in months.

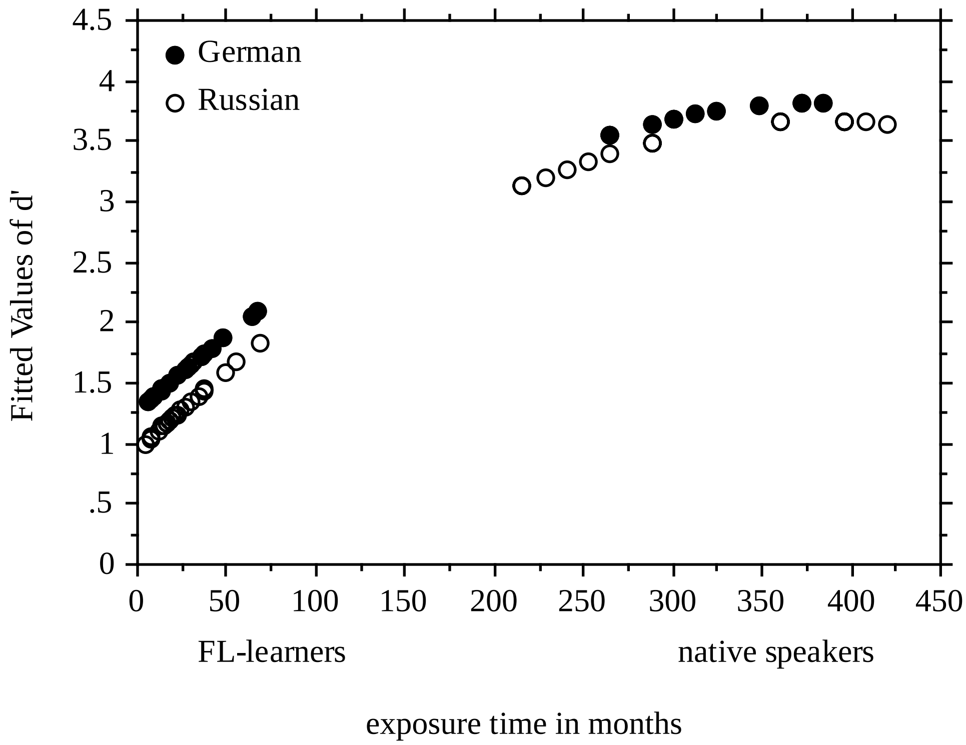

Figure 3: Quadratic fit of d' to exposure time for FL-learners and native speakers. Exposure time for native speakers is equal to their age in months.

The higher correlation of self-ratings with d-prime scores for the Russian learner group (.61) than for the German learner group (.39) may also reflect the decrease in the relative importance of lexical learning for the more advanced German language learners. Because of the problems Russian learners experience with the Cyrillic orthographic system, their overall level of lexical attainment after an equivalent period of exposure is not equal to that of the German learners. It appears that learners recognize the extent to which much of early language learning is based on rapid growth in vocabulary and that the correlations of d-prime with self-ratings reflect these judgments.

One way of explaining the slowed growth of lexical sensitivity at higher levels of exposure is that advanced learners begin to rely increasingly on lexical plausibility as opposed to actual semantic recognition. To test this account, we again calculated the chance-adjusted proportion of errors on pseudoderivatives among the false alarms. After this correction for chance, learners of German showed a .36 false alarm rate to pseudoderivatives, whereas learners of Russian showed a .01 false alarm rate (Scheffe's test, $p < .001$). This difference suggests that the learners of German rely more on lexical plausibility than the learners of Russian. This, however, did not affect the overall false alarm rates that were similar for both groups (.11 for German learners; .13 for Russian learners, $p > .5$).

A multiple regression analysis for each language group was performed to determine the impact of lexical plausibility on the progression of word sensitivity over time. In this analysis, as well as all remaining analyses, we are only looking at the two learner groups and not the native speaker groups. TotalST, errors on pseudoderivatives, and the interaction of these two measures were entered into the equation as predictor variables; the outcome variable was the d-prime score. The only significant effect for learners of Russian was TotalST ($F(1,18) = 7.7$; $p < .001$). However, in the German group, there was a positive effect for TotalST ($F(1,18) = 17.5$, $p < .001$) as well as for errors on pseudoderivatives ($F(1,18) = 15.6$, $p < .001$) and a significant negative effect for the interaction of TotalST with errors on pseudoderivatives ($F(1,18) = 12.3$, $p < .01$). This interaction indicates that advanced language learners show an increased negative impact of lexical plausibility on word sensitivity.

In order to examine in greater detail the nature of these effects for advanced learners, we transformed the exposure variables on a logarithmic scale, since logarithmic transformations allow one to assess curvilinear relations. Table 5 presents the correlations of the log-transformed exposure time variables with d-prime, overall mean reaction times, mean self-ratings, and CAPT scores. To understand the effects of the logarithmic transformation, please compare the first column of Table 4 with the first column of Table 5. The correlations of the time variables with d-prime in Table 4 are much weaker than the parallel correlations in Table 5 for the German FL group. This indicates that the increase in sensitivity over exposure time was much lower in the German FL group due to a greater impact of lexical plausibility on German learners as compared to Russian learners. At comparable levels of d-prime, Russian learners are not yet making full use of cues from Russian derivational morphology.

		d'	Overall RT	Self-Rating	CAPT
logFormalST	G	.39	-.09	.40	.22
	R	.46*	-.15	.43*	--
logInformST	G	.31	-.14	.22	.70**
	R	.53*	-.29	.54**	--
log TotalS	G	.44*	-.06	.39	.33
	R	.53*	-.29	.54**	--
log Breaks	G	.12	.23	.23	.07
	R	.21	-.11	.18	--
log Recency	G	.23	.44*	.06	.22
	R	.00	.69**	.31	--

** p < .01; * p < .05

Table 5: Correlations of dependent variables with log-transformed exposure time variables in the learner groups. For each variable designated in the left column the upper row (marked with G) represents the learners of German and the lower row (marked with R) the learners of Russian.

As noted earlier, Anderson and Freebody (1983) have argued that a yes/no test (like the lexical decision task) provides a more accurate measure of language proficiency than a multiple-choice test because it reduces the possibility of successful guessing. In particular, it could be the case that older subjects score higher on a multiple choice task like the CAPT because of increases in test taking experience which favor the development of successful guessing strategies. We tried to estimate the impact of test taking experience by entering the subject's age into the Multiple Regression analysis after all four exposure time variables. Indeed, we found a significant unique positive effect of age on the CAPT scores (F(1,16) =

4.9, $p < .05$) but not on d-prime or the self-ratings. This supports the analysis of Anderson and Freebody (1983) and indicates that the lexical decision task may have some advantages over multiple-choice proficiency tests in terms of providing a purer measure of underlying language attainment.

<u>D-prime and other exposure time variables</u>

Next, we wanted to determine whether periods of informal study time or immersion had more impact on the development of word sensitivity than periods of formal study. We also wanted to know whether breaks in regular exposure to the FL had an influence on word sensitivity.

The log-transformed time variables of interest were entered as predictors at subsequent steps into four sets of hierarchical regression analyses. The first analysis looked at the Russian learner group using d-prime as the criterion; the second conducted the same analylsis for learners of German. The third used the self-ratings score as the outcome variable for Russian learners; and the fourth was the same analysis for learners of German. In each of these analysis, we were interested in estimating how much variance is accounted for by each variable above and beyond Formal Study Time[4].

First, the effect of InformalST above and beyond FormalST on d-prime scores was examined. The unique effect related to immersion was nonsignificant for the learners of German ($p > .1$) but accounted for 23.1% of the variance for the learners of Russian ($F(1,19) = 7.9$, $p < .05$) with the sign of the regression coefficient being positive. Thus, immersion appeared to have a highly beneficiary effect for the learners of Russian in terms of increasing their d-prime sensitivity. A similar result was obtained for self-ratings as a outcome variable (GFL: $p > .3$; RFL: $F(1,19) = 8,2$; $p = .01$). Interestingly, for learners of German, InformalST had a significant positive effect on the CAPT scores ($F(1,19) = 18.9$; $p < .001$). Therefore, this variable was left in the equation.

Next, Breaks and Recency were entered in the equation in order to estimate the amount of decay in word sensitivity due to interruptions and the time passed since the learner's last regular exposure to the language. There were no significant unique effects of these variables on d-prime in either group, but there was a negative effect of Recency on the self-ratings for learners of Russian ($F(1,17) = 4.7$, $p < .05$). The more recent the language learning period, the higher the self-rating. But a gap of between one to three years between the end of language learning and the testing in this experiment, led to no signficant decrease in d-prime. This suggests that subjects seem to be able to maintain their state of lexical familiarity for a certain amount of time and that the lack of regular exposure to the language is more likely to reduce self-ratings, rather than actual lexical abilities.

Effects of exposure to other foreign languages

We next examined the effects for the learners of studying additional foreign languages. Table 6 shows the frequency, mean total study time, and mean self-ratings for the four major additional foreign languages, as well as all other languages combined.

Table 6

	Group	Frequency (number of subjects studying this language)	Total ST	Self-rating
French	German FL	11	55.1 (40.9)	3.2 (1.9)
	Russian FL	12	33.7 (30.0)	2.3 (1.9)
Spanish	German FL	3	21.0 (13.5)	2.4 (1.4)
	Russian FL	14	37.2 (30.8)	2.3 (1.6)
Latin	German FL	4	14.4 (9.3)	1.4 (0.5)
	Russian FL	3	27.7 (28.7)	1.2 (0.8)
Others	German FL	2	30.5 (41.7)	1.6 (1.2)
	Russian FL	11	19.8 (18.8)	1.8 (1.2)

Table 6: Study frequency, means, and standard deviations in total study time and self-ratings for additional FLs in the German and Russian learner groups.

Again, we conducted two separate stepwise regression analyses for each language -- one with d-prime as the outcome and one with self-ratings as the outcome. In addition, we conducted a fifth analysis for the German learner group with the CAPT scores as the outcome variable. The mean self-rating scores for all four languages were entered into the model as a block. If a given learner did not study a given language, the self-rating score was zero. This number is selected to indicate that there is no competition that derives from this particular source. As a whole, this block of variables did not yield significant unique effects either on d-prime nor on the self-ratings in the target language. However, if each language variable was entered separately there were marginally significant unique negative effects for Latin on d-prime in the German learner group ($F(1,16) = 4.0$; $p = .06$) and for French on d-prime in the Russian learner group ($F(1,16) = 3.6$; $p = .07$). Also, in the German learner group, there was a significant negative effect of "Other FLs" on the CAPT test scores ($F(1,16) = 6.7$, $p <.05$).

While there were no unique effects of each language on the Mean Ratings in the GFL group, there were significantly positive effects of Spanish ($F(1,16) = 5.0$; $p <.05$) and "Other FLs" ($F(1,16) = 11.2$; $p <.001$) in the RFL group. The data, therefore, suggest an inhibitory trend in word sensitivity caused by the study of additional foreign languages. However, these effects were highly selective with respect to the languages involved.

Language specific effects

Since we know that Russian and German differ considerably in the ease with which written input can be processed, we expect to find differences in language familiarity which are beyond the effects of exposure time or competing FLs. The main cause of differences between these two languages is the novelty of the graphemic system in Russian which takes longer to acquire and, thus, restricts the amount of input learners are able to process at the beginning stages of Russian learning.

To measure this effect, the model was fitted to the whole sample and Language (coded as a dummy variable with Russian as 0 and German as 1) was entered at the last step after the exposure time and additional foreign language study variables. Language accounted for an additional 11.5% of the variance of d-prime ($F(1,34) = 8.2$; $p <.01$) and 11.4% of the Mean Ratings ($F(1,34) = 6.8$; $p <.05$) which is due to overall higher sensitivity and higher self-rating in the German learner group. An additional check showed that there were no significant interactions of Language with the eight other predictor variables in the model.

Reaction times

Reaction times below 200 milliseconds were excluded, because it is not possible to make a true lexical judgment in such a brief period and such times must represent errors in response timing. Reaction times above 6000 milliseconds were also excluded, in accord with methodological suggestions presented by Ratcliff (1993). The use of these cut-off points led to the exclusion of 3.2% of the data points for the Russian learner group and 1.6% of the data points for the German learner group. Figure 4 presents the overall distributions of reaction times in the two learners groups. Although about 70% of the responses occur in the range below 2000 milliseconds, there is a significant number of responses at longer latencies. In general, these slower responses indicate a greater use of strategic processing in this subjects. The following analyses explore factors influencing the speed of lexical decisions in the Russian and German learners.

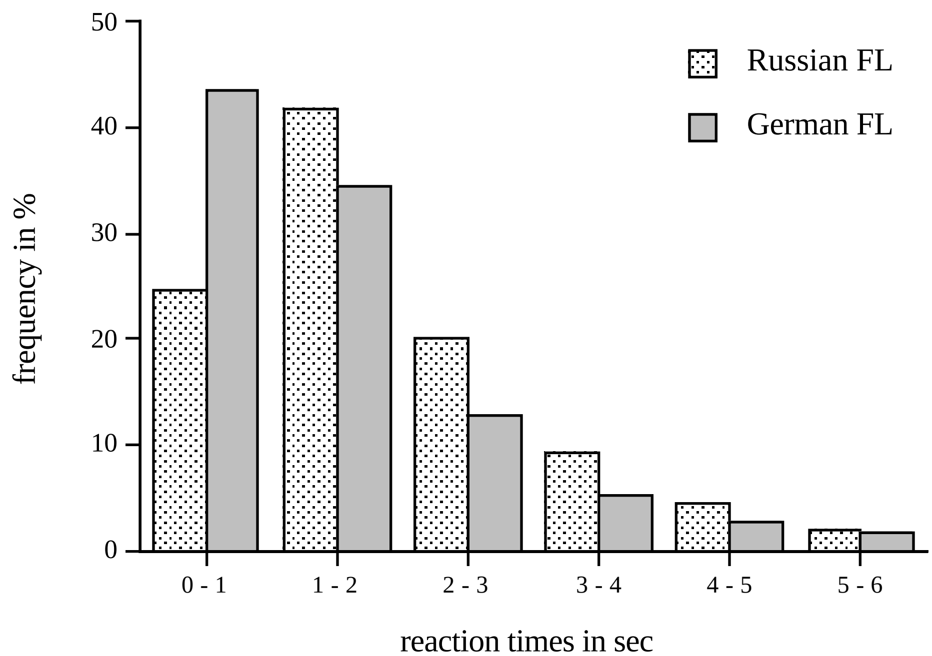

Figure 4: Overall frequency distribution of the reaction times in the lexical decision task for the learners of German and the learners of Russian

We conducted a simultaneous multiple regression analysis on the reaction times. The predictor variables in this analysis included the FormalST, InformalST, Breaks, Recency, Spanish study, French study, Latin study, and Other Language study. In this analysis, the joint and independent effects of the additional study variables were not significant in either group and were then removed. In the reduced model, the effects of the exposure time variables were similar. In both groups, there was a highly significant effect of Recency (German FL: $F(1,17) = 16.1$; $p = .001$; Russian FL: $F(1,17) = 23.2$; $p < .001$) with faster reaction times for learners who have been studying the language recently. Also, there was a significant effect of InformalST for the German learner group ($F(1,17) = 6.4$; $p < .01$) with faster reaction times for learners who had more immersion experience. This effect was only marginally significant in the Russian learner group ($F(1,17) = 3.5$; $p = .07$). Finally, Breaks had a significant effect in the German learner group ($F(1,17) = 9.6$; $p < .01$) with faster reaction times for learners who had less breaks in exposure.

The next analysis was performed to examine language differences in reaction times. The Language dummy variable, as well as all four interactions of Language with the exposure time variables were added to the model for the whole sample of language learners. While there were no significant interactions, Language yielded a strong independent effect ($F(1,38) = 25.3$; $p < .001$), indicating that the reaction times in the German learner group were reliably faster. The final model is given in Table 7. As can be seen, this model accounts for almost 60% of the total variance and shows a good fit.

	Overall RT		RT hit		RT miss	
	beta	t-value	beta	t-value	beta	t-value
Intercept	2395.0	10.017	1921.8	9.113	2612.4	8.184
Formal ST	-247.6	-1.457	-240.6	-1.606	-14.5	-.064
Informal ST	-104.7	-3.143**	-109.6	-3.731**	-80.7	1.815
Breaks	101.6	2.630*	81.5	2.393*	133.7	2.592*

Recency	182.5	5.925**	131.0	4.824**	242.1	5.888**
Language	-601.7	-5.036**	-304.8	-2.892**	-603.9	-3.782**
	R^2 = .592		R^2 = .494		R^2 = .522	
	adj R^2 = .538		adj R^2 = .427		adj R^2 = .459	
	model F (5,38)=11.0**		model F (5,38) = 7.4**		model F (5,38) = 8.3**	

Table 7: Simultaneous regression analysis for overall RTs, RTs for hits, and RTs for misses.

These analyses of overall reaction times do not tell us whether exposure time and decay affect just overall activation levels or whether there are specific effects related to response type. To disentangle these two possibilities, we performed the same analysis on the reaction times for hits and misses as dependent variables. Note that the reliability of the subject means varies as the proportions of hits and misses among the words vary. But since the highest hit vs. lowest miss rates did not exceed 20%, such an analysis seems to be justified given the large number of items. The minimum number of items involved in computing a given mean was 36.

These analyses showed that decreases in reaction times related to higher amounts of informal language study were entirely due to faster reaction times for hits, whereas the reaction times for misses were unaffected by increases in exposure time. The variables of Recency, Breaks, and Language, however, had significant effects on both correct and incorrect responses. This assymetry between the predictor variables in the analysis for separate response types serves to provide a possible explanation for the smaller magnitude of the exposure time variable compared to the decay variables in the analysis of the overall reaction times.

Discussion

The main findings of Experiment 2 are that the increase of word sensitivity over time was more rapid for learners of Russian, despite the fact that overall word sensitivity and reaction times were lower for Russian learners, once all other factors had been partialled out. We interpret this as evidence that the learners of German appeared to have reached a more advanced stage than the learners of Russian within the same amount of time, not only with respect to lexical knowledge but also with respect to their knowledge of orthographical and morphological patterns of the language. This leads, in turn, to a higher usage by the more advanced German learner group of lexical plausibility over semantic recognition in the lexical decision task. This foreshortening of the parts of the learning period that are most sensitive to lexical changes then led to smaller effects for the exposure variables for the German learners.

There was also an increase in the use of lexical plausibility for the presumably more advanced German learner group as well as the native speakers, as measured by false alarms to pseudoderivatives, although the level of lexical plausibility judgments among learners never quite reached native speaker levels (Russian difference = .278, Scheffe's test $p < .001$; German difference = .175, Scheffe's test $p < .05$). Despite the minor differences in the native speaker baseline, their increased false alarm rate for pseudoderivatives is easily interpreted as due to an increase of phonological, orthographic, and morphological knowledge during language learning.

One major difference between the two learner groups is the highly positive effect of immersion on d-prime and self-ratings for learners of Russian and the lack of this effect for learners of German. On the other hand, immersion had a highly positive effect on the CAPT scores for learners of German. A closer examination of the questionnaire data showed that the immersion exposures for the learners of Russian occurred mainly in the context of Student Exchange programs, whereas immersions for the learners of German occurred principally during private visits. This leads us to suspect that the intensity of immersion was less for the learners of German and was not sufficient to have an impact on their lexical familiarity or their self-ratings of language abilities. The high impact of immersion on the CAPT scores, however, is difficult to reconcile with this explanation, particularly, because the adaptive procedure and the random selection of test items for each subject make it difficult to specify the locus of the effect. In order to avoid unwarranted speculations, we would like to leave this question open, pending a deeper content analysis of the items on the CAPT.

The examination of effects of exposure to additional foreign languages demonstrated a trend towards inhibitory effects on lexical familiarity. An obvious explanation for this effect is that the concurrent study of several foreign languages can lead to a dilution in attentional resources during the learning process, thus resulting in weaker connections to the semantic representations in each language by adding more noise to the whole system. This is in accordance with multilingual interactive activation models (Grainger & Dijkstra, 1992; Keatley, Spinks & de Gelder, 1994) in that it suggests that the strength of intralingual connections needs to be developed slowly over time. If a student is exposed to both Russian and French vocabulary simultaneously, it will take some time to strengthen the links of each Russian word to the other Russian words and each French word to the other French words, so that these links will come to dominate over any potential links between languages or against background noise.

Interference from concurrent study of foreign languages is greatest when the two languages are of closely equal strength. This type of interlingual competition can be seen most clearly among those learners of Russian who had also studied French. This group had roughly similar exposure durations to the two languages and the correlation of self-ratings on the two languages was fairly high (+.34). This group is a major contributor to the finding that d-prime scores were lower in the Russian learner group. In comparison, the German learners who had also studied French typically had a much lower level of exposure to French and a negative correlation of self-ratings in the two languages (-.44). The marginal negative effect of Latin in the GFL group is striking, given the fact that the amount of Latin study time was generally small and similar in both groups. We can only speculate that the similarity between FLs also plays a role in that similar languages cause higher mutual inhibition than dissimilar languages. Latin is much closer to German than to Russian in terms of orthography and lexical roots. This closeness might have been responsible for the inhibitory effect.

The overall mean reaction times for the learners were generally below that of native speakers, indicating a generally weaker development of lexical automaticity. Decreases in reaction times related to higher amounts of informal language study were entirely due to faster reaction times for hits whereas the reaction times for misses were unaffected by increases in exposure time. The variables of Recency, Breaks, and Language, however, had significant effects on <u>both</u> correct and incorrect responses.

The important result here is that reaction times become faster with higher levels of exposure and slower when there are breaks or interruptions in that exposure. However, the fact that we did not find an effect of exposure time on the negative responses seems to suggest that learners with who are a bit “rusty” at the language adopt a strategy of allowing additional search time if a positive response is not triggered within a critical amount of time. By allowing this additional search time, these learners are able to maintain a d-prime level that is not decremented for the effects of gaps in exposure. We can view this as evidence that lack of exposure causes established connections and access pathways to weaken but not to vanish completely.

Conclusions

The results of these two studies indicate that we have identified a new assessment instrument that can play an important role in the study of foreign language learning. This new tool has five important positive characteristics.

1. <u>Systematic test construction procedure</u>. The construction of the test materials for Russian and German was based on the systematic application of a replicable

frequency-based method for item construction. The lexical decision performance of native speakers in Experiment 1 indicated that this algorithm was successful in obtaining fairly equidistant word and nonword distributions in Russian and German. However, minor differences in the results for native speakers in the two languages indicate that further standardization of the test construction procedure could lead to further improvements.

2. Crosslinguistic baseline. It is possible to devise comparable versions of this task in languages with markedly different lexical structures. This ability to devise parallel versions for different languages can provide us with the backbone for crosslinguistic studies of lexical processing in foreign language learning. This technique can be applied to any language for which reliable frequency counts are available. We are currently aware of frequency accounts for nearly 30 languages. However, as tools for the computerized processing of text data become more generally accessible, this number will continue to grow and our ability to construct additional versions of this test will also grow.
3. Crosslinguistic differences. The word length and pseudoderivative differences identified in Experiment 1 suggest that we might be able to use this instrument in future studies as a way of examining how variations in the target language could affect the process of lexical learning. In Experiment 2, the novelty of the Cyrillic graphemic system led to inferior lexical decision performance for learners of Russian. These two results show that the lexical decision task is sensitive to language specific differences in learning speed which cannot be captured by exposure time. This task provides an alternative method for directly comparing learners of different languages.
4. Psychometric validities. The high correlations of the lexical decision task with other measures of language attainment, such as CAPT proficiency scores and self-ratings indicate that the measure has good construct validity and predictive validity.

Moreover, one could argue that the strong psycholinguistic grounding of the lexical decision task gives it a higher construct validity than comparable conventional measures. The fact that we found a significant unique positive effect of age on the CAPT scores, but not on d-prime indicates that the lexical decision task may be less sensitive to test-taking strategies and may provide a purer measure of underlying language attainment.

5. Convenience of administration. An important aspect of this measure is the fact that it can be administered in a fully consistent manner within a space of ten to fifteen minutes and scored by the computer in seconds. Moreover, the examiner does not need to receive any special training in test administration or scoring.
6. Linkage to exposure variables. Apart from these important methodological characteristics, the results of these two studies underscore the extent to which the both the d-prime and reaction time measures are tightly coupled to increases in language exposure, as well as interruptions in exposure and competition from other foreign languages.

We would like to suggest that these six positive attributes of the lexical decision task argue strongly for its use as a basis for the crosslinguistic study of the initial periods of language learning. In particular, we recommend that researchers use d-prime scores derived from this measure as the principle grouping variable or covariate in future studies of foreign language learning.

Construction of tests of this type provide us with just a beginning wedge in the development of a more effective approach to crosslinguistic studies of foreign language learning. Ideally, we would like to have a complete armamentarium of experimental techniques for measuring each of the component processes involved in speaking, reading, writing, and listening to a foreign language (MacWhinney, in press). We would like to measure component processes involved in processing phonology, orthography, prosody, lexicon, syntax, and discourse. Even within the area of lexical processing, it is clear that much more work can be done. We will need to supplement a visual lexical decision task of this type with other tasks such as auditory lexical decision or naming (Simpson & Krueger, 1991), picture-name priming (Schriefers, Meyer & Levelt, 1990), and lexical processing during sentence interpretation (MacWhinney & Bates, 1989; MacWhinney, Bates & Kliegl, 1984; McDonald & MacWhinney, 1995). A fuller understanding of crosslinguistic patterns in language learning will not emerge until we have secured a much greater control over the basic methodology of crosslinguistic assessment.

References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. Advances in Reading Language Research, 2, 231-256.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database. Philadelphia: Linguistic Data Consortium.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. Journal of Experimental Psychology: Human Perception and Performance, 10, 340-357.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task. Journal of Memory and Language, 24, 89-106.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel distributed processing approaches. Psychological Review, 100, 589-608.

Feldman, L. B., & Turvey, M. T. (1983). Word recognition in Serbo-Croation is phonologically analytic. Journal of Experimental Psychology: Human Perception and Performance, 9(2), 288-298.

Flege, J., & Davidian, R. (1984). Transfer and developmental processes in adult foreign language speech production. Applied Psycholinguistics, 5, 323-347.

Fodor, J. (1983). The modularity of mind: An essay on faculty psychology. Cambridge, Mass.: M. I. T. Press.

Forster, K. (1976). Accessing the mental lexicon. In R. Wales & E. Walker (Eds.), New approaches to language mechanisms . New York: North Holland Publishing Company.

Forster, K. I. (1985). Lexical acquisition and the modular lexicon. Language and Cognitive Processes, 1, 87-108.

Friederici, A., Weissenborn, J., & Kail, M. (1991). Pronoun comprehension in aphasia: A comparison of three languages. Journal of Memory and Language, 1-21(29), 336-360.

Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. Journal of Experimental Psychology, 20, 116-129.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. Journal of Experiment Psychology: Human Perception nd Performance, 13, 104-115.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. Journal of Verbal Learning and Verbal Behavior, 22, 22-44.

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. Harris (Ed.), Cognitive processing in bilinguals . Amsterdam: North-Holland.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. Cognitive Psychology, 21, 60-99.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: the status of subjacency in the acquisition of a second language. Cognition, 39, 215-258.

Keatley, C., Spinks, J., & de Gelder, B. (1994). Asymmetical cross-language priming effects. Memory and Cognition, 22, 70-84.

Keller, H. H. (1991). Word frequency and pedagogical value: contrasting textbook vocabulary lists with word frequency counts. Slavic and East European Journal, 35, 228-244.

Kilborn, K., & Ito, T. (1989). Sentence processing in Japanese-English and Dutch-English bilinguals. In B. MacWhinney & E. Bates (Eds.), The crosslinguistic study of sentence processing . New York: Cambridge University Press.

Lukatela, G., Popadic, D., Ognjenovic, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. Memory and Cognition, 8, 124-132.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. Harris (Ed.), Cognitive processing in bilinguals . Amsterdam: Elsevier.

MacWhinney, B. (in press). Language specific prediction in foreign language acquisition. Language Testing, in press, xxx-xxx.

MacWhinney, B., & Bates, E. (Eds.). (1989). The crosslinguistic study of sentence processing. New York: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. Journal of Verbal Learning and Verbal Behavior, 23, 127-150.

McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. Applied Psycholinguistics, 8, 379-414.

McDonald, J. L., & MacWhinney, B. J. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. Journal of Memory and Language, 34, 543-566.

Meara, P., & Buxton, B. (1987). An alternative to multiple-choice vocabulary tests. Language Testing, 4, 142-154.

Paradis, M., & Lebben, G. (1987). The assessment of bilingual aphasia. Hillsdale, NJ: Lawrence Erlbaum.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological Bulletin, 114, 510-532.

Schriefers, H., Meyer, A., & Levelt, W. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language, 29, 86-102.

Sebastián-Galles, N. (1991). Reading by analogy in a shallow orthography. Journal of Experimental Psychology: Human Perception and Performance, 17, 471-477.

Seidenberg, M., & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Simpson, G., & Krueger, M. (1991). Selective access of homograph meanings in sentence context. Journal of Memory and Language, 30, 627-643.

Stone, G. (1994). Top-down and bottom-up phonology in visual word recognition : Psychonomics Society Meeting, St. Louis.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. Journal of Experimental Psychology: Human Perception and Performance, 19, 744-774.

Swinney, D. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 645-659.

Walley, A. C. (1988). Spoken word recognition by young children and adults. Cognitive Development, 3, 137-165.

White, L. (1990). The verb-movement parameter in second language acquisition. Language Acquisition, 4, 337-360.

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive negative evidence in the classroom. Second Language Research, 7, 133-161.

Wulfeck, B., Bates, E., Juarez, L., Opie, M., Friederici, A., MacWhinney, B., & Zurif, E. (1989). Pragmatics in aphasia: Crosslinguistic evidence. Language and Speech, 32, 315-336.

Zasorina, L. N. E. (1977). Castotnyj slovar' russkogo jazyka. Moscow: Russkij jazyk.

Appendix 1: German and Russian stimulus material listed in ascending word frequency

German:

Abnutzung - Bewagnung
Düne - Lühme
Fessel - Mepfer
Erstling - Ballheit
Geiz - Leum
anfragen - umwachen
Begierde - Entschelte
biegen - rieden
Hecke - Saspe
Segnung - Wagnung
Fremdheit - Klinkheit
Gelöbnis - Ersollnis
Geiger - Belter
Volkstum - Glaubtum
Feigheit - Stadtheit
erachten - bepfannen
Tarnung - Machung
Eber - Toger
Gelübde - Bereude
Nabel - Fubel
feilen - käulen
Paarung - Leibung
Windung - Sonnung
Bücherei - Kleiderei
Erbarmen - Gewohle
Pfütze - Malpe
grimmig - wärmig
spärlich - menglich
putzen - mipfen
Büchse - Mockse
angeln - wurteln
stammeln - respeln
Senf - Glupf
plagen - tuhmen
Sohle - Pfiele
Besoldung - Verhabung
Schraube - Bähse
Rückweg - Herfeld
Schmutz - Kips
nähen - walden
Grundriß - Hochschnitt
dehnen - rahlen
Schranke - Sempe
heulen - käunen
Pforte - Wörke
Neuling - Kurzling
Dose - Nube
brummen - krungen
Wettlauf - Topfgang
Piste - Luspe
Raub - Spreil
pochen - pfecken
Puppe - Kruffe
brüllen - germen
Vorort - Nachplatz
Kragen - Stuhnen
feierlich - hilflich
gleiten - dauten
Scherz - Gülm
büßen - pöhsen
Obst - Dehst
klettern - hittern
Strich - Sproll
eilen - euben
Eifer - Auler
Reiz - Golz
Festung - Breitung
Vieh - Tohe
leugnen - nurgen
Zunge - Bunne
Brille - Linger
loben - ruben
irren - sampen
Bescheid - Ersitz
Anblick - Belauf
Muster - Asper
neigen - däuben
schildern - mulkern
Teppich - Ropper
Anklage - Entsinge
lenken - burken
bewirken - entsitzen
Einfuhr - Ausgrund
brennen - spallen
Brot - Niehm
Schmerz - Vost
Unrecht - Beklang
Tasche - Goche
Keller - Käpper
weinen - läumen
Vorhaben - Umreden
wieviel - woland
Schuld - Pfalt
Sicherung - Stielung
empfangen - zubilden
Stoff - Nolk
Schicht - Blark
bereiten - ertreten
Pferd - Dalk
Rücken - Pfocken
fremder - bonder
Dasein - Zuhang
beenden - verdeuten
begreifen - umbitten
zahlen - wielen
Angriff - Aufplatz
Mond - Schlapf
Hof - Buht
bauen - weuen
Bruder - Dehder
lernen - olden
Glück - Pfoch
Tisch - Gilf
kaufen - scheiken
laut - burm
Mitglied - Zuraum
langsam - tiefsam
Ordnung - Klebung
lesen - geien
Tür - Blör
Brief - Krohm
endlich - breitlich
Wert - Tild
übrige - aufige
Besuch - Verfund
handeln - fonneln
Schule - Nehle
gern - piehl
bedeuten - erhelfen
Kirche - Durke
schwerer - tober
zehn - zwahn
Meinung - Kämpfung
neben - soden
Ziel - Soht
lange - rolde
wichtig - schönig
allein - unmal
erhalten - entlingen
Monat - Dehser
Betrieb - Anbund
mal - gohl
Ende - Ralde
Woche - Bische
zwar - lor
welcher - weres
Weg - Vod
Haus - Keut
Teil - Nien
heißen - hargen
sprechen - gechsen
kleiner - weuner
Staat - Deel
Leben - Soben
wo - wuh
dürfen - tilpen
wissen - zunnen
Frau - Kreu
Uhr - Ool
Mann - Wepf
nun - dam
Land - Sulm
Tag - Terp
heute - ulten
sehr - bling
Welt - Biel
jeder - dager
erster - dreite
Mensch - Klüff
gehen - gahlen
hier - vubst
uns - seus
doch - juch
gegen - ögen
schon - sähm
geben - bieren
neuer - heier
wollen - darren
sagen - meumen
dann - gull
sollen - mätten

Russian:

oznob - obig
vymå - lemå
girå - madå
vrednosth - ploxosth
ovod - ivol
popirathså - okivathså
perepad - prodar
mleth - xreth
desna - garza
oslik - rebik
ukrowenie - naglawenie
r˛avçina - platçina
ßalfej - vargej
velikodußie - golomyslie
obolhwenie - opiwenie
tykva - vokra
krewenie - zvonenie
˛u˛˛ath - nurzith
zagon - somin
webetath - trevlewath
repa - terå
vovleçenie - izywenie
vra˛debnosth - verxosth
gadanie - ugnilenie
˛ivosth - bystrosth
maket - izlog
utolåth - ulith
rasporoth - preslepith
beremennaå - iznytyj
˛rebij - zdorij
kuvyrkom - otlopom
kutathså - globithså
obnimka - dovërtka
grim - tlan
÷elithså - utrithså
razval - peregrëb
stog - glik
miloserdie - zlotvorie
kulëk - bruvik
burlith - strometh
krasnoreçie - uzkoçuvstvie
måth - çlåth
pastbiwe - begbiwe
razduvath - presopeth
soxa - urma
izvinenie - perezarenie
sosulhka - edalka
xrapeth - glovath
noçleg - poloxod
muravej - krapogej
otklik - perevid
uni˛ath - otvyßath
testo - preßko
otbor - otglåd
mstith - ˛leth

pervyj - dvotyj
celovek - gologod
novyj - vertyj
kakoj - pekoj
delo - gumo
u˛e - ave

lysina - zråçina
zaçastu√ - otskorå
lazith - braçith
napitok - iz˛arok
musor - grasel
laåth - niåth
pugathså - zåtithså
veslo - ßysto
darith - novith
lu˛a - pawa
pepel - votel
mudrosth - dru˛esth
nalog - izgar
greth - kleth
ovowi - lywi
orex - istax
leçith - podreth
opozdath - uzigath
igrußka - stoåßka
izdatelh - çistitelh
kot - z√th
soedinåth - otzidath
nagnuthså - izmythså
owuwenie - ußnosth
kryßka - furka
myth - steth
privykath - zavetith
myßlenie - ruçenie
åbloko - morono
razbudith - sovitath
korolh - bagolh
spasenie - tonenie
pid˛ak - grobin
muxa - ra˛a
torgovath - ßorovith
dyxanie - dalenie
platith - dyleth
urok - arzëm
doçka - vaçka
prodath - iztrith
påtno - gl√no
krovath - taronh
gotovithså - golovathså
moloko - zirino
zvuk - glåm
stroith - brazith
bolhno - ˛alhno
osenh - uradh
√˛nyj - me˛ij
izuçath - izgreçath
˛elanie - pozovie
ozero - krizalo
vraç - trunh
nazyvathså - ubolethså
pti÷a - drë˛a
zakryth - izdnith
uxo - reso
spina - strulhva

kupith - ëmith
liçnyj - werbyj
napravlenie - preserdie
naçathså - izynåthså
deti - zaty
˛elath - vagath
ugol - orin
rost - nårs
gotovyj - zemåtyj
gosth - vrosth
zelënyj - kra˛uçij
sçët - glot
poexath - posvoith
çuvstvo - rivstvo
desåth - prycth
iskath - ostith
rebënok - ve˛olok
tå˛ëlyj - treglonnyj
davno - çatro
÷elh - gralh
skolhko - drutko
råd - n√r
okolo - vlåmo
uli÷a - lobi÷a
reßith - wånath
çitath - gribath
okno - √rmo
myslh - modh
bystro - mustro
segodnå - wednå
doroga - kiloga
prijti - trejti
protiv - u˛ov
kone÷ - wune÷
va˛nyj - kormyj
puth - vleth
trud - drolh
znacith - ˛ricith
ponimath - izmykath
pocemu - pocego
gorod - ˛orobh
cerez - prå
togda - çeda
raboçij - mucij
dath - zith
ka˛dyj - edyj
slovo - mlevo
strana - gråga
dol˛en - tu˛en
zdesh - bridh
raz - goz
xoteth - dåveth
denh - monh
ocenh - abedh
nado - ilo
˛iznh - graznh

bolhßoj - kortoj
znath - neth
sebå - kalå
da - be
naß - dus
govorith - zelobith

Appendix 2: Language textbooks used for estimating frequency of occurence in FL learners

German:

Byrnes, H., Fink, S. (1987) Wendepunkt: Intermediate German for Proficiency. Heinle & Heinle.

Crean, J. E., Scott, M., Briggs, J. (1993) Deutsche Sprache und Landeskunde. New York: McGraw-Hill.

Lohnes, W. F.W, Strothmann, F.W., Petig, W.E. (1989) German: A structural approach. New York: Norton.

Moeller, J., Liedloff, H., Adolph, W. R., Kirmse, C., Lalande, J. F. (1992) Deutsch heute: Grundstufe. Boston: Houghton Mifflin.

Terell, T.D., Tschirner, E., Nikolai, B., Genzmer, H. Kontakte: A communicative approach. New York, McGraw-Hill.

Russian:

Bitekhtina, G., Davidson, D., Dorofeyeva, T., Fedyanina, N. (1988) Russian Stage One. Moscow: Russky Yazyk Publ.

Leed, R., Nakhimovsky, A., Nakhimovsky, A, (1981) Beginning Russian, Vol. 2, Columbus, OH: Slavica.

Leed, R., Nakhimovsky, A., Nakhimovsky, A, (1982) Beginning Russian, Vol. 1 , Columbus, OH: Slavica Publishers Inc.

Stepanova, E. M., Ievlev, Z. N., Trushina, L. B., Baker, R. L. (1984) Russian for Everybody, Moscow: Russky Yazyk Publ.

Townsend, Ch. E. (1981) Continuing with Russian, Columbus, OH: Slavica.

---

[1] Note that stress is sometimes inconsistent in Russian suggesting that some lexical involvement is required.

[2] However, the clustering of graphemes related to one single phoneme in German introduces the possibility that the letter is simply not the relevant unit for phonological assembly of German words. To check this we coded the number of phonemes for each word in both languages and substituted word length with number of phonemes in the ANCOVA. In Russian, the number of phonemes and letters is largely identical; differences occur only due to the presence of the soft sign which palatalizes the preceeding consonant. In German, the

---

following grapheme clusters each correspond to one single phoneme: *sch*, *ch*, *ck*, *ie*, all double consonants, all double vowels, and vowels followed by *h*. We also coded diphtongs as single phonemes. Consequently, the correlation between number of letters and number of phonemes for our stimuli is higher in Russian (.96) than in German (.90). The ANCOVA with number of phonemes as covariate yielded marginally significant main effects of Language ($F(1,706) = 5.48$, $p < .05$) and Lexicality ($F(1,706) = 3.71$; $p = .05$), as well as a highly significant effect of the Phonemes covariate ($F(1,706) = 95.0$, $p < .001$). The tests for common slopes of the regression lines on different levels of the three factors were highly significant for the 2-way interactions of Phonemes x Language ($F(1,706) = 13.1$, $p < .001$) and Phonemes x Lexicality ($F(1,706) = 14.0$, $p < .001$) and marginally signficant for the 3-way interactions of Phonemes, Lexicality and Derivation ($F(1,706) = 4.7$, $p < .05$) and Phonemes x Language x Lexicality ($F(1,706) = 4.3$, $p < .05$). No other effects reached significance. This model accounted for 72.3 % of the variance although word frequency again was nor included The substitution of letters by phonemes did not improve the model fit neither did it lessen the Language x Length interaction which indicates that the word length effect, be it operationalized by number of letters or number of phonemes, is much more pronounced in Russian.

[3] The inclusion of a general language proficiency test for the estimation of external validity is problematic since input based language familiarity and proficiency in various areas might be quite independent. Language proficiency tests are based on the assumption that learning in different language domains progresses more or less in parallel. However, besides higher susceptibility to individual differences, other domains like morphology or syntax are also subject to language specific effects which might impose very different demands on the learner. In fact, one of our goals in assessing language familiarity is to construct a measure which language differences in the development of other domains can be related to. Therefore, correlations with the CAPT will be considered only as a very rough indicator for external validity.

[4] In order to avoid missing values in the logarithmic transformations 0 months of Informal study time, Breaks, and Recency were coded as .001.