

Video Article

Lexical Decision Task for Studying Written Word Recognition in Adults with and without Dementia or Mild Cognitive Impairment

Alexandre Nikolaev¹, Eve Higby², JungMoon Hyun³, Sameer Ashaie⁴¹Helsinki Collegium for Advanced Studies, University of Helsinki²University of California, Riverside³Hunter College, City University of New York⁴Feinberg School of Medicine, Northwestern UniversityCorrespondence to: Alexandre Nikolaev at alexandre.nikolaev@helsinki.fiURL: <https://www.jove.com/video/59753>DOI: [doi:10.3791/59753](https://doi.org/10.3791/59753)

Keywords: Word recognition, Lexical decision, Psycholinguistics, Aging, Dementia, Alzheimer's disease, Mild Cognitive Impairment, Mixed-effects models, Principal components analysis

Date Published: 6/19/2019

Citation: Nikolaev, A., Higby, E., Hyun, J., Ashaie, S. Lexical Decision Task for Studying Written Word Recognition in Adults with and without Dementia or Mild Cognitive Impairment. *J. Vis. Exp.* (), e59753, doi:10.3791/59753 (2019).

Abstract

Older adults are slower at recognizing visual objects than younger adults. The same is true for recognizing that a letter string is a real word. People with Alzheimer's disease (AD) or Mild Cognitive Impairment (MCI) demonstrate even longer responses in written word recognition than elderly controls. Despite the general tendency towards slower recognition in aging and neurocognitive disorders, certain characteristics of words influence word recognition speed regardless of age or neuropathology (e.g., a word's frequency of use). We present here a protocol for examining the influence of lexical characteristics on word recognition response times in a simple lexical decision experiment administered to younger and older adults and people with MCI or AD. In this experiment, participants are asked to decide as quickly and accurately as possible whether a given letter string is an actual word or not. We also describe mixed-effects models and principal components analysis that can be used to detect the influence of different types of lexical variables or individual characteristics of participants on word recognition speed.

Video Link

The video component of this article can be found at <https://www.jove.com/video/59753/>

Introduction

Words are stored in the mental lexicon in a highly interconnected network. The connections between words may reflect shared properties, such as semantic similarity (e.g., *dog* and *cat*), form similarity (*dog* and *fog*), or frequent co-occurrence in common language use (e.g., *dog* and *leash*). Cognitive theories of language, such as usage-based theory¹, argue that every encounter of a word by a language user has an effect on the word's mental representation. According to Exemplar Theory, a word's representation consists of many exemplars, which are built up from individual tokens of language use and which represent the variability that exists for a given category. The frequency of use² impacts representations in memory by contributing to the strength of an exemplar¹.

Word recognition speed can reveal the characteristics of the mental lexicon. A commonly used experimental paradigm for measuring the speed of word recognition is the lexical decision task. In this task, participants are presented with letter strings on a monitor, one at a time. They are instructed to decide as quickly as possible whether the letter string on the screen is a real word or not by pressing the corresponding button.

By examining reaction times for real words, researchers can address a number of important questions about language processing. For example, identifying which factors make recognition faster can test hypotheses about the structure of the mental lexicon and reveal its architecture. Moreover, comparisons of performance across different groups of participants can help us understand the influence of various types of language experience, or, in the case of aging or neurodegenerative diseases (e.g., Alzheimer's disease), the role of cognitive decline.

Some factors (e.g., the frequency of use) exhibit greater influence on word recognition than other factors (e.g., word length). With advancing age, the way people recognize written words might change^{3,4}. Younger adults tend to rely heavily on semantic (meaning-based) aspects of a word, such as how many compounds (e.g., *bulldog*) or derived words (e.g., *doggy*) share aspects of both form and meaning with the target word (in this case, *dog*). Word recognition for older adults appears to be more influenced by form-based aspects, such as the frequency that two subsequent letters co-occur in the language (e.g., the letter combination *st* occurs more often in English words than the combination *sk*).

To determine the factors that influence the word recognition speed across different groups, the researcher can manipulate certain variables in the stimulus set and then test the power of these variables to predict word recognition speed. For example, to test whether word recognition is driven by semantic or form-based factors, the stimulus set should include variables that reflect the degree of connectivity of a word to its semantic neighbors in the mental lexicon or its connectivity to other words that share part of its form.

This method was used in the current study to investigate whether word recognition speed is influenced by different factors in younger and older adults and in individuals with Alzheimer's disease (AD) or mild cognitive impairment (MCI)³. The method described here is based on visual word recognition but can be adapted to the auditory modality. However, some variables that are significant predictors of reaction times in a typical visual lexical decision experiment might not predict response latencies in an auditory lexical decision or may have the opposite effect. For example, the phonological neighborhood has the opposite effect across these two modalities⁵: words with larger phonological neighborhoods exhibit a facilitatory effect on visual word recognition but result in longer response latencies in auditory lexical decision⁶.

Word-finding difficulties in older adults⁷ have been generally attributed to difficulty accessing the phonological word form rather than a breakdown of the semantic representation⁸. However, AD research has primarily focused on semantic declines^{9,10,11,12,13,14}. It is important to disentangle how semantic and orthographic factors influence the recognition of written words in aging with and without cognitive decline. The influence of form-related factors is more pronounced in older than in younger adults, and it remains significant in people with MCI or AD³. Thus, this methodology can help us uncover features of the mental lexicon across different populations and identify changes in the lexicon's organization with age and neuropathology. One concern when testing patients with neuropathology is that they may have difficulties accessing task-related knowledge. However, the lexical decision task is a simple task with no burden on working memory or other complex cognitive skills that many patients exhibit problems with. It has been considered appropriate for AD and MCI populations.

Protocol

The protocol follows the guidelines of the Ethics Committee of the Hospital District of Northern Savo (IRB00006251).

1. Participant screening

1. Recruit younger and older adults who have normal or corrected-to-normal vision and are native speakers of the language tested unless the study addresses specific research questions regarding second language acquisition.
2. For healthy control groups, exclude participants who have a history of neurological or psychiatric disorders.
3. For the clinical groups, recruit individuals who have been diagnosed with Alzheimer's disease¹⁵ or mild cognitive impairment^{16,17}. Recruit only individuals who are able to give informed consent, according to the clinician's judgment. For accurate comparisons, match the age range and mean of the clinical groups with that of the healthy older adult participants.
4. Measure the severity of dementia, for example, using the Clinical Dementia Rating Scale¹⁸ (CDR, 0=no dementia, 0.5=very mild, 1=mild, 2=moderate, 3=severe). Exclude patients with severe dementia because the task may be too difficult for them. Do not include participants who seem unable to follow instructions, despite their severity rating.

2. Stimulus construction

1. Select word stimuli to address specific research questions, for example, whether semantic or orthographic/phonological variables have a stronger influence on word recognition¹⁹ in different populations.
2. Calculate from a corpus²⁰ or retrieve from a database²¹ variables reflecting semantic, phonological, and orthographic characteristics of the stimuli so they can be used either as theoretically motivated predictors explaining word recognition reaction times or as control variables. Also, use participants' gender, age, and years of education as explanatory or control variables.
3. In addition to the real words, build a set of matched pseudo-words. Pseudo-words resemble real words in that they conform to the language's norms for placement of certain letters in certain word positions (phonotactics). In order to control for phonotactics, create pseudo-words, for example, by randomly recombining the first syllables from some words with the second syllables from other words. Remove any items that happened to produce a real word through this recombination and all the items that violate the phonotactics of the language.
4. Match the pseudo-words with the target words in terms of the word length in letters and bigram frequency, which is the average number of times that all combinations of two subsequent letters occur in a text corpus. These variables have been shown to influence recognition speed. NOTE: Manipulating the pseudo-word ratio (e.g., the number of real words relative to the number of pseudo-words) may lead to different results, with responses to the less probable stimuli being slower and less accurate²².
5. Add a set of real-word fillers in order to decrease participant's expectancy of the next stimulus belonging to a certain type (e.g., a certain inflectional class). Choose them, for instance, from different word categories (e.g., inflectional classes) than the ones used to construct stimuli according to the characteristics of interest.

3. Experimental design

1. Present letter strings horizontally, one at a time, subtending a visual angle of about 5°.
2. Begin the experiment with a practice session that includes a small number of trials, with one word presented per trial (e.g., 15 words and 15 pseudo-words not included in the actual experiment). This is to familiarize the participant with the task and the response buttons. If the participant is not responding accurately ('yes' button for real words and 'no' button for pseudo-words) during the practice trial, provide feedback and redo the practice session.
3. Divide the experiment into blocks and give short breaks after the practice session and between the blocks. These breaks allow participants to rest their eyes and will reduce fatigue.
4. Start each new block with a few filler items that will not be included in the analysis (e.g., common nouns such as *dog*, *sister*, *year*) because the first few trials of the block are sometimes ignored by participants with MCI or AD.
5. Present the experimental items in a random order for each participant.
6. Begin each trial with a fixation mark (e.g., a + sign) appearing in the center of the screen for 500 ms, followed by a blank screen for a fixed (e.g., 500 ms) or variable amount of time (e.g., 500-800 ms).
7. Immediately after the blank screen, present a letter string (word or pseudo-word) for 1,500 ms or until the participant responds.

8. After a response is made or after 1,500 ms from the onset of the word (whichever comes first), follow again with a blank screen until 3000 ms has passed from the beginning of the trial.
 9. Repeat this sequence until all of the items in the experiment have been presented.
- NOTE: Times for the delay between the stimuli serve as an example. Changing them may affect the pattern of results.

4. Experimental procedure

1. Place the participant in front of a computer monitor at a viewing distance of about 80 cm in a normally lit room.
 2. Instruct the participant to decide as quickly and accurately as possible whether the letter string on the screen is a real word or not by pressing one of two corresponding buttons with their dominant hand (e.g., the index finger for real words and the middle finger for pseudo-words) or using the index finger of each hand.
- NOTE: Participants try to optimize their performance in line with the instructions. Thus, their responses will be affected by stressing speed over accuracy or vice versa²³.

5. Analyzing data with a mixed-effects model in R

NOTE: Many different statistical programs can be used to perform the analysis. This section describes steps for analyzing data in R²⁴.

1. Obtain the reaction time (RT) measured in milliseconds for each trial from the output file of the presentation program (e.g., E-Studio software).
2. Install the packages *lme4*²⁸ and *lmerTest*²⁹. Attach packages with the function *library* or *require*.
3. Import data into R by using, e.g., the *read.table* function.
4. Check the need for transformation, e.g., with the *boxcox* function from the *MASS* package²⁵, as the distribution of RT data is typically highly skewed.


```
> library(MASS)
> boxcox(RT ~ Explanatory_variable, data = yourdata)
```

NOTE: The graph produced by the *boxcox* function shows a 95% confidence interval for the boxcox transformation parameter. Depending on the lambda values located within this interval, the needed transformation can be chosen, e.g., $\lambda = -1$ (inverse transformation), $\lambda = 0$ (logarithmic transformation), $\lambda = 1/2$ (square root transformation), and $\lambda = 1/3$ (cube root transformation).

 1. Transform the RT values using inverted transformed RTs (e.g., $-1000/RT$) or binary logarithms of RTs (e.g., $\log_2(RT)$) since these transformations tend to provide more normal-like distributions for reaction times in lexical decision experiments than raw RTs²⁶.
 2. Alternatively, use statistical methods that do not rely on normal distributions and fit robust linear mixed-effects models and provide estimates on which outliers or other sources of contamination have little influence²⁷.
5. Since reaction time analyses are typically conducted on accurate responses, exclude trials in which the participants' response was incorrect (a response of "no" to real words) as well as omissions.
 1. Also, exclude responses to pseudo-words and fillers unless there are specific hypotheses about them.
 2. Exclude trials with response times faster than 300 ms because they typically indicate that the participant was too late responding to a previous stimulus or that he or she accidentally pressed the response button before reading the stimulus.
6. Build a basic linear mixed-effects model that identifies *RT* as the outcome measure and *Subject*, *Item*, and *Trial* as random effects. Note that variables whose values are randomly sampled from a larger set (population) of values are included as random effects and variables with a small number of levels or for which all levels are included in the data are fixed effects. Add the random effects in the form $(1 | Subject)$ in order to estimate random intercepts for each of the random effects.


```
> g1 = lmer(RT ~ (1 | Subject) + (1 | Item) + (1 | Trial), data = yourdata)
> summary(g1)
```
7. Add explanatory variables in a theoretically motivated order. For instance, add words' base frequency as a fixed effect. Some variables, such as base or surface frequency, have Zipfian distributions, so insert them in the model with a transformation that results in a more Gaussian distribution shape, e.g., logarithmic transformation.


```
> g2 = lmer(RT ~ log(BaseFrequency + 1) + (1 | Subject) + (1 | Item) + (1 | Trial), data = yourdata)
> summary(g2)
```
8. Check with the *Anova* function if adding each predictor (e.g., *BaseFrequency*) significantly improved the predictive power of the model compared to a model without the predictor.


```
> anova(g1, g2)
```

 1. If there is no significant difference in the fit of the new model over the simpler model, prefer the simplest model with fewer predictors. Also, check the Akaike Information Criterion (AIC)³⁰ of each model. AIC is a measure of how well statistical models fit a set of data according to maximum likelihood. Lower values indicate a better fit for the data³¹.


```
> AIC(g1); AIC(g2)
```
9. Repeat steps 5.7. and 5.8. by adding other explanatory variables, e.g., some of those that are presented in **Table 1**, one by one in a theoretically motivated order and keeping only those that significantly improve the predictive power of the model. If variable stimulus onset asynchrony was used, include it as a fixed-effect variable in the model.
10. Check for theoretically motivated interactions between predictors. For instance, add a term of interaction *the log of Base Frequency by Age*.


```
> g3 = lmer(RT ~ log(BaseFrequency + 1) + Age + log(BaseFrequency + 1) : Age + (1 | Subject) + (1 | Item) + (1 | Trial), data = yourdata)
```

NOTE: It is possible that a predictor is significant as a term of interaction with another variable, but not significant as the main predictor. In this case, do not remove this predictor from the model (include it also as the main effect).
11. Add by-participant random slopes³² for predictors by including "1 +" before the variable name, then "| Subject", e.g., $(1 + \log(BaseFrequency + 1) | Subject)$, because participants' response times might be affected by words' lexical characteristics in different ways.

NOTE: If there are many continuous predictors, allowing them all to have random slopes is unrealistic because random slope models require large amounts of data to accurately estimate variances and covariances^{33,34}. In case the maximal model does not converge (in other words, successfully compute), simplify the model³³. Alternatively, implement Bayesian versions of multilevel modeling³⁵.

12. Run the analysis for each participant group separately. Alternatively, run an analysis on all data, with group as a fixed-effect predictor, and then test for an interaction of group by significant predictors.
`> g4 = lmer (RT ~ log(BaseFrequency + 1) + Age + log(BaseFrequency + 1) : Age + Group + log(BaseFrequency + 1) : Group + (1 + log(BaseFrequency + 1) | Subject) + (1 | Item) + (1 | Trial), data = yourdata)`
13. In order to remove the influence of possible outliers, exclude data points with absolute standardized residuals exceeding, e.g., 2.5 standard deviations²⁶, and re-fit the model with the new data (*yourdata2*).
`> yourdata2 = yourdata [abs(scale(resid(g4))) < 2.5,]`
`> g5 = lmer (RT ~ log(BaseFrequency + 1) + Age + log(BaseFrequency + 1) : Age + Group + log(BaseFrequency + 1) : Group + (1 + log(BaseFrequency + 1) | Subject) + (1 | Item) + (1 | Trial), data = yourdata2)`

NOTE: Not all extreme data points are harmful for the model – only those that have excessive leverage over the model.
14. In the case of exploratory (data-driven) analysis, use backward stepwise regression: include all variables in the initial analysis and then remove non-significant variables from the model in a step-by-step fashion. Alternatively, use the automatic procedure of eliminating non-significant predictors with the *step* function provided by the package *lmerTest*²⁹.
`> step (g4)`

Representative Results

Table 1 shows a list of variables that were obtained from three different sources (a corpus, a dictionary, and pilot testing of test items) that are included in the analysis as fixed-effect predictors. Many of these variables have been previously reported to affect word recognition speed.

Corpus:	
Base frequency	the number of times a word appears in the corpus in all its different forms (e.g., child and children)
Bigram frequency	the average number of times that all combinations of two subsequent letters occur in the corpus
Morphological family size	the number of derived and compound words that share a morpheme with the noun
Morphological family frequency	the summed base frequency of all morphological family members
Pseudo-morphological family size	includes not only “true” morphological family members but also words that mimic morphological family members in their orthographic form, whether or not they are actual morphemes, and thus represents orthographic overlap but not necessarily semantic overlap
Pseudo-morphological family frequency	the summed base frequency of all pseudo-morphological family members
Surface frequency	the number of times a word appears in the corpus in exactly the same form (e.g. child).
Trigram frequency	the average number of times that all combinations of three subsequent letters occurs in the corpus
Dictionary:	
Hamming distance of one	the number of words of the same length but differing only in any single letter ³⁶
Length	number of letters
Orthographic neighborhood density	the number of words with the same length but differing only in the initial letter ^{37,38}
Pilot testing:	Sixteen participants indicated on a six-point scale (from 0 to 5) their estimates for each of the target words on the following parameters.
As proper name	how often the word is seen as a proper name (e.g., as a family name, like Baker) ³⁹
Concreteness	the directness with which words refer to concrete entities ⁴⁰
Familiarity rating	how familiar the word is
Imageability	the ease and speed with which words elicit mental images ⁴⁰

Table 1. The variables included in the mixed-effects analysis as fixed-effect predictors, obtained from three different sources (a corpus, a dictionary, and pilot testing of test items).

The number of explanatory variables can be smaller or bigger depending on the research questions and on the availability of the variables from databases, dictionaries, or corpora. However, including a large number of lexical features as predictors might lead to complications in the form of collinearity between predictors, when predictors correlate with each other and thus exert similar effects on the outcome measure. For example, concreteness and imageability of words may be highly correlated. An assumption in any linear regression analysis is that the predictor variables are independent of each other. However, as more variables are added to the model, the risk that some of the variables are not independent of each other increases. The higher the correlation between the variables, the more harmful this collinearity can be for the model⁴¹. A potential consequence of collinearity is that the significance level of some predictors may be spurious.

To avoid the effect of collinearity between predictors, the number of predictors should be reduced. If two predictors show collinearity, only one of them should be included in the model. However, if more than two predictors show collinearity, then excluding all but one would lead to a loss of variance explained. On the one hand, a researcher might reduce the number of explanatory variables already in the experimental design *a priori*, leaving only those that are hypothesis driven (theoretically motivated) and that permit the researcher to test hypotheses between different populations. On the other hand, sometimes there is no existing theory, and thus, it is reasonable to use Principal Component Analysis (PCA)⁴¹ to reduce the number of predictors by combining predictors that have similar effects into components. In this analysis, the predictor space was orthogonalized and the principal components of the new space were used as predictors (following steps described here⁴¹ on pages 118-126). One disadvantage of using PCA is that sometimes the components make it difficult to disentangle the effects of multiple predictors; they might all emerge with strong loadings on the same principal component.

We transformed all lexical predictors into five principal components to examine how word recognition speed might be different for younger adults and older adults. However, only two of them were significant in the young adults' data (Table 3): PC1 and PC4. Three principal components (PCs) were significant predictors in the model for elderly controls (Table 4), MCI (Table 5) and individuals with AD (Table 6).

	PC2
Bigram freq.	-0.390
Hamming distance of one	-0.350
Final trigram freq.	-0.330
Neighborhood density	-0.320
Length	-0.226
Initial trigram freq.	-0.224
Pseudo-family size (final)	-0.124
Pseudo-family freq.(final)	-0.052
Family freq. (compounds)	-0.042
Family size (compounds)	-0.039
Family freq. (derived words)	-0.036
Family size (derived words)	-0.034
Surface freq.	-0.023
Base freq.	-0.008
Pseudo-family size (initial)	0.070
Familiarity rating	0.093
As proper name	0.102
Pseudo-family freq. (initial)	0.113
Concreteness	0.275
Imageability	0.296
Pseudo-family size (internal)	0.296
Pseudo-family freq. (internal)	0.316

Table 2. The rotation matrix for PC2. The loadings are the degree to which each variable contributes to the component. This table has been modified with permission from Cortex³.

Table 2 presents the lexical variables with their loadings on PC2. The strongest positive loadings of PC2 were pseudo-family size and frequency for overlap in the internal position. The strongest negative loadings were bigram frequency, Hamming distance of one, final trigram frequency, and orthographic neighborhood density. Since all of these variables are primarily form-based rather than meaning-based, PC2 is interpreted as reflecting the influence of form-based aspects of a word on word recognition speed.

Tables 3 shows the results of the mixed-effects analysis for young adults (31 participants). Since PC2 was not a significant predictor of young adults' response times (see Table 3), this seems to indicate that these form-based variables have less influence on the young adults' reaction times compared to older adults', including those with AD or MCI.

Fixed effects	Estimate	Std.Error	t-value	p-value
(Intercept)	-1.31	0.05	-26.36	<0.001
Allomorphs	-0.034	0.015	-2.3	0.024
PC1	-0.021	0.004	-5.179	<0.001
PC4	-0.042	0.008	-5.224	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Item	(Intercept)	0.009	0.095	
Subject	(Intercept)	0.032	0.179	
	PC1	4.765e-05	0.007	0.08
Residual		0.005	0.235	
Number of obs. 2862; Item, 99; Subject, 31				

Table 3. Estimated coefficients, standard errors, and t- and p-values for the mixed models fitted to the response latencies elicited for real words for young adults. This table has been modified with permission from Cortex³.

The *Estimate* for a fixed-effect variable can be interpreted as the amount by which the dependent variable (RT) increases or decreases if the value of this fixed effect changes. If the *Estimate* is negative, it means the variable correlates negatively with reaction times (the higher the variable, the smaller (faster) the reaction times). The *t-value* should typically be less than -2 or greater than 2 in order for the predictor to be significant.

Table 4, Table 5, and Table 6 show the results of the mixed-effects analysis for elderly controls (17 participants), individuals with MCI (24 participants), and individuals with AD (21 participants).

One interesting difference between the three elderly groups emerged: education significantly predicted speed of word recognition in elderly controls (**Table 4**; the estimate for Education is negative, which means that more years of education was associated with faster reaction times) and individuals with MCI (**Table 5**), but not in individuals with AD (**Table 6**; Education was dropped from the model since it was not a significant predictor), although there was no obvious difference in the variability of years of education among these groups (AD: mean 10.8 years, SD = 4.2, range = 5-19; MCI: mean 10.4 years, SD 3.5, range 6-17; elderly controls: mean 13.7 years, SD 3.7, range 8-20).

Fixed effects	Estimate	Std.Error	t-value	p-value
(Intercept)	-0.72	0.157	-4.574	<0.001
Allomorphs	-0.022	0.01	-2.14	0.035
PC1	-0.011	0.003	-4.122	<0.001
PC2	-0.011	0.005	-2.223	0.029
PC4	-0.02	0.006	-3.687	<0.001
Education	-0.024	0.011	-2.237	0.041
Random effects				
Groups	Name	Variance	Std.Dev.	
Item	(Intercept)	0.003	0.057	
Subject	(Intercept)	0.026	0.16	
Residual		0.033	0.181	
Number of obs. 1595; Item, 99; Subject, 17				

Table 4. Estimated coefficients, standard errors, and t- and p-values for the mixed models fitted to the response latencies elicited for real words for elderly controls. This table has been modified with permission from Cortex³.

Fixed effects	Estimate	Std.Error	t-value	p-value
(Intercept)	-0.562	0.114	-4.922	<0.001
PC1	-0.009	0.003	-3.218	0.002
PC2	-0.013	0.005	-2.643	0.01
PC4	-0.018	0.006	-3.078	0.003
Education	-0.039	0.01	-3.708	0.001
Random effects				
Groups	Name	Variance	Std.Dev.	
Item	(Intercept)	0.003	0.056	
Subject	(Intercept)	0.03	0.174	
Residual		0.061	0.248	
Number of obs. 2227; Item, 99; Subject, 24				

Table 5. Estimated coefficients, standard errors, and t- and p-values for the mixed models fitted to the response latencies elicited for real words for individuals with MCI. This table has been modified with permission from Cortex³.

Fixed effects	Estimate	Std.Error	t-value	p-value
(Intercept)	-0.876	0.051	-17.017	<0.001
Allomorphs	-0.018	0.009	-2.008	0.048
PC1	-0.011	0.003	-4.097	<0.001
PC2	-0.011	0.004	-2.718	0.008
PC4	-0.018	0.005	-3.751	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Trial	(Intercept)	0.001	0.034	
Item	(Intercept)	0.002	0.049	
Subject	(Intercept)	0.045	0.212	
	PC1	4.138e-05	0.006	0.83
Residual		0.026	0.162	
Number of obs. 1879; Item, 99; Subject, 21				

Table 6. Estimated coefficients, standard errors, and t- and p-values for the mixed models fitted to the response latencies elicited for real words for individuals with AD. This table has been modified with permission from Cortex³.

The study reported here addressed an additional question: whether the number of stem allomorphs associated with a word influences the speed of word recognition^{42,43}. Stem allomorphs are different forms of a word stem across various linguistic contexts. For example, in English, *foot* has two stem allomorphs, *foot* and *feet*. In other words, the word stem changes depending on whether it is in the singular or plural form. The study described here tested speakers of Finnish, a language that has quite a bit more complexity in its stem changes compared to English. Words with greater stem allomorphy (i.e., words with more changes to their stems) elicited faster reaction times in all groups (Table 3, Table 4, and Table 6; the estimates for the number of allomorphs were negative, which means the higher the number of allomorphs a word had, the faster the reaction times it elicited) except the MCI group (Table 5; the number of allomorphs was not a significant predictor and hence was dropped from the model).

Discussion

By using a simple language task that does not require language production, the present study investigated the impact of various lexical variables on word recognition in neurologically healthy younger and older adults, as well as in people with Alzheimer’s disease or Mild Cognitive Impairment. The age range used for recruiting “older adults” might depend on the specific research interests; however, the range for the healthy elderly group should match as closely as possible the age range and distribution for individuals with MCI or AD recruited for the same study.

To avoid collinearity between predictors, the lexical variables were orthogonalized into principal components and added to the mixed-effects models, where reaction times served as the dependent variable. The combination of a simple lexical decision experiment and a mixed-effects regression analysis led to the novel finding that the language difficulties for patients with AD may be attributed not only to changes to the semantic system but also to an increased reliance on word form. Interestingly, a similar pattern was found for people with Mild Cognitive Impairment and cognitively healthy elderly. This suggests that an increased reliance on form-based aspects of language processing might be part of a common age-related change in written word recognition.

In a factorial design, researchers traditionally create two or more sets of words that differ according to the variable of interest and then match these sets of words on a number of other lexical characteristics that may influence processing speed. The assumption is that any behavioral difference obtained between these two sets of words should be attributed to the manipulated (i.e., unmatched) variable. One problem with this type of design is that it is very difficult to match sets of words on more than a few variables. Another problem is that there might be some potentially significant variables that the word sets were not matched on or could not be matched on for a variety of reasons. Also, the factorial design treats continuous phenomena as if they are dichotomous factors. The use of mixed-effects models for statistical analysis of the behavioral data permits the researcher to include potentially important lexical variables as explanatory variables without the need to match words or lists of words according to these variables. In a mixed-effects model the variables *Subject* (participant code/number), *Item* (experimental stimuli), and *Trial* (trial number) are added as random effects. The random intercepts were included because it is assumed that subjects vary in their overall reaction times (i.e., some participants are naturally slower or faster across the board)

This methodology can be applied to other types of questions and to other populations, e.g., multilinguals or individuals with aphasia. For the latter group, language processing may differ from monolinguals, so this variable should be considered if recruiting a mixed-language population, either by restricting recruitment to only one type of group or by comparing results later to determine whether language background influenced results.

Disclosures

The authors have nothing to disclose.

Acknowledgments

We thank Minna Lehtonen, Tuomo Hänninen, Merja Hallikainen, and Hilikka Soininen for their contribution to the data collection and processing reported here. The data collection was supported by VPH Dementia Research enabled by EU, Grant agreement No. 601055.

References

1. Bybee, J. From Usage to Grammar: The Mind's Response to Repetition. *Language*. **82** (4), 711-733 (2006).
2. Oldfield, R.C., Wingfield, A. Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology*. **17**, 273-281 (1965).
3. Nikolaev, A., et al. Effects of morphological family on word recognition in normal aging, mild cognitive impairment, and Alzheimer's disease. *Cortex*. **116**, 91-103 (2019).
4. Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., Baayen, R. H. Discrimination in lexical decision. *PLoS ONE*. **12** (2), e0171935, 1-42 (2017).
5. Andrews, S. Frequency and neighborhood size effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **15**, 802-814 (1989).
6. Grainger, J., Muneaux, M., Farioli, F., Ziegler, J.C. Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology*. **58A** (6), 981-998 (2005).
7. Ossher, L., Flegal, K. E., Lustig, C. Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition*. **20**, 220-242 (2013).
8. Barresi, B. A., Nicholas, M., Connor, L. T., Obler, L. K., Albert, M. Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*. **7**, 169-178 (2000).
9. Chertkow, H., Whatmough, C., Saumier, D., Duong, A. Cognitive neuroscience studies of semantic memory in Alzheimer's disease. *Progress in Brain Research*. **169**, 393-407 (2008).
10. Cuetos, F., Arce, N., Martínez, C. Word recognition in Alzheimer's disease: Effects of semantic degeneration. *Journal of Neuropsychology*. **11**, 26-39 (2015).
11. Stilwell, B. L., Dow, R. M., Lamers, C., Woods, R. T. Language changes in bilingual individuals with Alzheimer's disease. *International Journal of Language & Communication Disorders*. **51**, 113-127 (2016).
12. Obler, L. K. Language and brain dysfunction in dementia. In S. Segalowitz (Ed.), *Language functions and brain organization*. (pp. 267-282). New York, NY: Academic Press (1983).
13. Obler, L. K., Albert, M. L. Language in the elderly aphasic and in the demented patient. In M. T. Sarno, (ed.) *Acquired aphasia*. (pp. 385-398). New York: Academic Press (1981).
14. Obler, L. K., Gjerlow, K. *Language and the brain*. Cambridge: Cambridge University Press (1999).
15. McKhann, G. M., et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging - Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dementia*. **7**, 263-269 (2011).
16. Winblad, B., et al. Mild cognitive impairment - beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*. **256**, 240-246 (2004).
17. Albert, M. S., et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging - Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. **7**, 270-279 (2011).
18. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., Martin, R. L. A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*. **140**, 566-572 (1982).
19. Baayen, R. H. Data Mining at the Intersection of Psychology and Linguistics. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*. (pp. 69-83). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers (2005).
20. Brants, T., Franz, A. *Web 1T 5-gram, version 1*. Philadelphia: Linguistic Data Consortium (2006).
21. Baayen, R. H., Piepenbrock, R., Gulikers, L. *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium (1995).
22. Wagenmakers, E. J., Ratcliff, R., Gomez, P., McKoon, G. A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*. **58**, 140-159 (2008).

23. Dufau, S., Grainger, J., Ziegler, J. C. How to say "no" to a non-word: a leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **38**, 1117-1128 (2012).
24. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2018).
25. Venables, W. N., Ripley, B. D. *Modern applied statistics with S*. (4th ed.). New York, NY: Springer (2002).
26. Baayen, R. H., Milin, P. Analyzing reaction times. *International Journal of Psychological Research*. **3** (2), 12-28 (2010).
27. Koller, M. robustlmm: An R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*. **75** (6), 1-24 (2016).
28. Bates, D., Mächler, M., Bolker, B., Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. **67**, 1-48 (2015).
29. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*. **82**, 1-26 (2017).
30. Akaike, H. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.), *Second International Symposium on Information Theory*, (pp. 267-281). Academiai Kiado: Budapest (1973).
31. Sakamoto, Y., Ishiguro, M., Kitagawa, G. *Akaike Information Criterion Statistics*. D. Reidel Publishing Company (1986).
32. Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. **68**, 255-278 (2013).
33. Bates, D., Kliegl, R., Vasishth, S., Baayen, H. Parsimonious mixed models. *arXiv:1506.04967v2* (2015).
34. Harrison, X.A., et al. Brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*. **6**, e4794, 1-32 (2018).
35. Kimball, A.E., Shantz, K., Eager, C., Roy, C.E.J. Confronting quasi-separation in logistic mixed effects for linguistic data: a Bayesian approach. *Journal of Quantitative Linguistics*. (2018).
36. Coltheart, M., Davelaar, E., Jonasson, J. T., Besner, D. Access to the internal lexicon. In S. Dornick, (ed.) *Attention and performance, vol. VI*. (pp. 535-556). Erlbaum, Hillsdale, New Jersey (1977).
37. Caselli, N.K., Caselli, M.K., Cohen-Goldberg, A.M. Inflected words in production: Evidence for a morphologically rich lexicon. *The Quarterly Journal of Experimental Psychology*. **69**, 432-454 (2016).
38. Yarkoni, T., Balota, D., Yap, M. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*. **15** (5), 971-979 (2008).
39. Cohen, G. Recognition and retrieval of proper names: Age differences in the fan effect. *European Journal of Cognitive Psychology*. **2** (3), 193-204 (1990).
40. Kemmerer, D. *Cognitive neuroscience of language*. New York: Psychology Press (2015).
41. Baayen, R.H. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press (2008).
42. Nikolaev, A., Lehtonen, M., Higby, E., Hyun, J., Ashaie, S. A facilitatory effect of rich stem allomorphy but not inflectional productivity on single-word recognition. *Applied Psycholinguistics*. **39**, 1221-1238 (2018).
43. Nikolaev, A., et al. Behavioural and ERP effects of paradigm complexity on visual word recognition. *Language, Cognition and Neuroscience*. **10**, 1295-1310 (2014).