Experimental Study on a Two-string Lexical Decision Task: Non-Words and Words

Shuyu Zhang
The Australian National University, Canberra, Australia

Sihong Zhang (✉)
Hefei University of Technology, Hefei, China
zhangsihong@hfut.edu.cn

Abstract—This paper introduces the general purposes, hypotheses and designs of the lexical decision task and compares the results of several existing studies. Based on previous studies, three hypotheses are proposed. Then, it illustrates a two-lexical decision task designed and completed by the Research School of Psychology, Australian National University. In comparison with traditional lexical decision task, the two-string lexical decision task further tests participants' response time to non-words and words. The results of the current two-string lexical decision task experiment verify the validity of previous studies on the one hand, while on the other hand, do not fully support the statement that participants would make faster responses to unrelated words than unrelated non-words. The findings of the current study directly provide cognitive processes for English lexical differentiation and learning, which could give hints to English lexical teaching and acquisition.

Keywords—Lexical decision task, non-words, words, response time

1 Introduction

The lexical decision task has been widely used for studying visual word recognition over the last two decades. Lexical decisions based on a continuous source of word-likeness evidence for both non-words and words [1]. The task involves clarifying letter strings either as words or as non-words, and it requires participants to react as quickly as possible without making errors. Letter strings in words obey the rules of spelling for the given language and those in non-words do not obey the rules of spelling for that language [1]. Studies have investigated the impact of word-likeness on longer word strings presented concurrently and in different combinations. This paper considers word and non-word strings two words in length that is these letter strings are presented in pairs. It has been argued that the two-dimensional signal detection framework, that is pairs of words referred to in the literature as two-string, indicates discriminability adequately [2]. Unlike the traditional lexical decision task, a two-string lexical decision task requires participants to decide whether both strings

are words, or one string is made up of non-words [3]. Several studies [4, 5, 6] investigated the response time (RT, hereafter) and accuracy of participants under different types of word strings. Some researchers have found that participants tended to make a slower and less accurate response while reacting to the non-word stimuli that looked like words [1]. Others have argued it took longer for participants to reject pseudowords sharing phonemic similarity with words [7].

Wagenmakers, et. al [4] suggested that the RT and accuracy rate were highly related to the lexical characteristics of the experimental stimuli in their study of a typical lexical decision task. For example, participants generally responded more slowly and less accurately to the non-words that look like words than the non-words that do not look like words. According to the diffusion model account of lexical decision tasks [2], the term *drift rate* refers to the rate with which information is accumulated and it depends on the familiarity of the information. Drift rates are highest for high-frequency words, followed by low-frequency words, very-low frequency words, pseudowords and random letter strings. It is suggested that the visual similarity is strongly related to the drift rate. Wagenmakers et al.'s study showed that participants in this typical lexical decision task required more time to respond to the non-words sharing similar visual characteristics than words sharing similar visual characteristics and meanwhile make more mistakes [4].

Shulman, Hornak and Sanders [8] compared the RT and accuracy rate of different word pairs with different characteristics. They investigated word pairs like BRIBE-TRIBE, which shared similarity both graphically and phonetically. They also studied word pairs like FREAK-BREAK, which were graphically similar but dissimilar phonetically. In contrast, participants responded to word pairs like BRIBE-TRIBE more quickly and accurately than word pairs like FREAK and BREAK. In this study, the hypothesis was that the mode of lexical access is not phonetically mediated. Consequently, the task of determining whether a pseudoword is a non-word does not involve phonetic coding. The searching process activated the stimulus in short-term memory and directed queries into lexical memory. The slow RT of word pairs like FREAK and BREAK was caused by processing the incorrect pronunciation of the second word.

The present study aims to compare two basic conditions of forms of word or non-word pairs, pairs with visual similarity (e.g. JUGDE sounds like JUDGE) and pairs with phonetic similarity (e.g. KEAP sounds like KEEP) on accuracy effect and RT. The previous studies have laid the foundation for studying different word string types, however, *the direct comparison between visually and phonetically similar word strings is so far not sufficient*, which is also the present study aiming to testify. From the previous studies, it was predicted that the lexical characteristics of word strings and the process of activating the short-term memory are influential to the RT and accuracy. We thus hypothesised that:

- Participants would make a faster and more accurate response on a word string with a non-word that sounds like a word than a word string comprised of a non-word visually similar to the word;

- Participants would respond more slowly and less accurately to both types of non-words than related words
- Participants would make faster responses to unrelated words than unrelated non-words.

2 Method

2.1 Participants

Experiment 1 "Looks Like Word": The 166 participants in this group were undergraduates majoring in psychology from the Research School of Psychology, Australian National University (ANU), consisting of 70 males and 96 females, with ages ranging from 17 to 29 years old, and a mean age of 19.33 (SD = 1.99).

One hundred and forty-five of the participants were right-handed, 17 of them, left-handed and another four were ambidextrous. One hundred and twenty-three of the participants had English as their first language. Although the others did not have English as their first language, they could speak fluent English.

Experiment 2 "Sounds Like Word": The 154 participants in this group were also undergraduate psychology students from the same school of the ANU, consisting of 109 males and 45 females, with ages ranging from 18 to 25 years, and a mean age of 19.06 (SD = 1.37). One hundred and forty-three of the participants were right-handed, eight of them, left-handed and another three were ambidextrous. One hundred and nineteen of the participants had English as their first language and the other students could use English as well as their first language.

Participants were randomly assigned in Experiment 1 and Experiment 2.

2.2 Material

The 84-item lexical decision task consisted of 36 WORD/WORD pairs and 48 WORD/NON-WORD pairs. The word pairs in the task were designed to measure RT and accuracy in seven different conditions: strongly-related words, weakly-related words, unrelated words, filler words, strongly-related non-words, weakly-related non-words, and unrelated non-words. "Strongly-related" WORD/NON-WORD pair includes a non-word member of the pair that either looks like (Experiment 1) or sounds like (Experiment 2) a particular word that is strongly-related to the word member of the pair, for example, BUTTER vs. BRAED. "Weakly-related" WORD/NON-WORD pair includes a non-word member of the pair that either looks like (Experiment 1) or sounds like (Experiment 2) a particular word that is weakly-related to the word member of the pair, for example, CYCLE vs. TRAYN. "Unrelated" WORD/NON-WORD pair includes non-word member that looks like (Experiment 1) or sounds like (Experiment 2) a particular word that is not related to the word member of the pair, for example, HAMMER vs. FOME.

A demographic data sheet that asked for information such as age, gender, first language was also included.

2.3   Procedure

Two experiments ran during the same one experimental session with two different test groups: Group of "Looks Like Word" (Experiment 1) and Group of "Sounds Like Word" (Experiment 2). The participants were divided into two groups by seats and each participant took part in one of the two experiments in a random way. Before the experiment, participants filled in their information in the demographic data sheet. They were then instructed to complete the task on the lab computers with the following criteria: trying to respond as quickly and accurately as they could; making sure that eyes were paralleled to the middle of the computer screen; sitting approximately 57cm to the computer screen and responding to the words by clicking the button. Before the formal task, participants were given a practice session to understand the rules. During the task, the word pairs appeared randomly on the screen. Each word pair stays at the screen for 400ms. A 1-minute break was given to participants in the middle of the task. After finishing the task, participants were asked to record their own RT and accuracy on the statistic sheet with the given information on the screen. We collected the answer sheets from students at the end of the experiment and then analysized the data by using SPSS version 23.0 to calculate their mean reaction time, standard deviations and p value ($p < .05$ indicates statistical differences).

3   Results

After the experiments, we collected 115 effective answer sheets that met the experimental requirements from Experiment 1. Sheets from 11 males and 40 females were excluded, respectively. From Experiment 2, we collected 111 effective answer sheets with 29 sheets from males and 13 sheets from females excluded, respectively. The exclusion criteria include participants that were over the age of 30 years; that have not completed data sets; that got fewer than eight out of twelve correct in the "Both Words" Conditions; that gave too long RT in more than three of the seven experimental conditions.

An initial analysis of the data found significant differences between the RT of participants toward word strings and non-word strings in both Experiment 1: Looks like Word and Experiment 2: Sounds like Word.

The first hypothesis was that participants' RT would be faster while responding to non-words that sound like words than while responding to non-words that look like words. In other words, average RT in non-words conditions in Experiment 2 would be faster than that in Experiment 1. In the condition of Strongly-Related Non-words, the results of a paired samples *t*-test provided support for this prediction, $t(225) = 4.334$, $p < .001$. Participants demonstrated faster reaction time while responding to non-words that sound like words (M = 795.01, SD = 125.21) than while responding to non-words that look like words (M = 871.58, SD = 140.31). In the condition of Weakly-Related Non-words, the results were $t(225) = 5.331$, $p < .001$. Participants made shorter RT while responding to non-words that sound like words (M = 786.22, SD = 126.36) than while responding to non-words that look like words (M = 886.02, SD = 153.94). In the condition of Unrelated Non-Words, the results were $t(225) = 1.592$, $p = .013$. Par-

ticipants respond faster to non-words that sound like words (M = 785.35, SD = 123.74) than non-words that look like words (M = 813.13, SD = 138.51). See Table 1, Table 2 and Table 3 below for more information.

Table 1. Summary of Mean Reaction Time and Standard Deviations (in milliseconds) for Condition in the Lexical Decision Task – Experiment 1: Looks Like Word (*N* = 114)

Condition	Mean	SD
Strongly-Related Words	727.80	105.88
Weakly-Related Words	766.32	125.91
Unrelated Words	803.38	139.48
Strongly-Related Non-Words	871.58	140.31
Weakly-Related Non-Words	886.02	153.94
Unrelated Non-Words	813.13	138.51
Filler Condition	784.68	124.02

Table 2. Summary of Mean Reaction Time and Standard Deviations (in milliseconds) for Condition in the Lexical Decision Task – Experiment 2: Sounds Like Word (N = 111)

Condition	Mean	SD
Strongly-Related Words	727.27	114.87
Weakly-Related Words	772.03	132.77
Unrelated Words	810.60	146.71
Strongly-Related Non-Words	795.01	125.21
Weakly-Related Non-Words	786.22	126.36
Unrelated Non-Words	785.35	123.74
Filler Condition	766.11	117.57

Table 3. Between-Subject Analysis for Non-Word Conditions

Conditions	T-statistic
Strongly-Related Non-Words	$t(225) = 4.334, p < .001$
Weakly-Related Non-Words	$t(225) = 5.331, p < .001$
Unrelated Non-Words	$t(225) = 1.592, p = .013$

The second hypothesis was that participants would require more time to related non-words than related words in both experiments. In other words, all the data in the two experiments relevant to Related Words vs. Related Non-words should be statistically significant, that is, $p < .05$. This was the case as shown in Table 4 and Table 5 below with all the data of this kind demonstrating the value of $p < .05$. For example, the participants' RT of Strongly-Related Words and Weakly-Related Words should be faster than that of Strongly-Related Non-Words and Weakly-Related Non-words, respectively. In the condition of Strongly-Related Words vs. Strongly-Related Non-words, the results of paired samples *t*-test supported this, $t(114) = -18.852, p < .001$ in Experiment 1. In Experiment 2, the results were $t(111) = -9.694, p < .001$. In the condition of Weakly-Related Words vs. Weakly-Related Non-words, the results of paired samples *t*-test also provided supports for this hypothesis, $t(114) = -12.756, p < .001$ in Experiment 1. In Experiment 2, the results were $t(111) = 2.868, p = .005$.

Table 4. Paired-Samples T-Test for Experiment 1

(1)   3 Word Conditions and 3 Non-Word Conditions

Conditions	T-statistic
Strongly-Related Words vs. Strongly-Related Non-Words	$t(114) = -18.852, p < .001$
Weakly-Related Words vs. Weakly-Related Non-Words	$t(114) = -12.756, p < .001$
Unrelated Words vs. Unrelated Non-Words	$t(114) = -1.065, p = .289$
Strongly-Related Words vs. Weakly-Related Words	$t(114) = -6.583, p < .001$
Strongly-Related Words vs. Unrelated Words	$t(114) = -9.435, p < .001$
Weakly-Related Words vs. Weakly-Related Non-Words	$t(114) = -4.442, p < .001$
Strongly-Related Non-Words vs. Weakly Related Non-Words	$t(114) = -1.988, p = .049$
Strongly-Related Non-Words vs. Unrelated Non-Words	$t(114) = 7.648, p < .001$
Weakly Related Non-Words vs. Unrelated Non-Words	$t(114) = 8.694, p < .001$

(2)   4 Non-Word Conditions

Conditions	T-Statistic
Strongly-Related Non-Words vs. Weakly-Related Non-Words	$t(114) = -1.988, p = .049$
Strongly-Related Non-Words vs. Unrelated Non-Words	$t(114) = 7.648, p < .001$
Strongly-Related Non-Words vs. Filler(Baseline) Condition	$t(114) = 12.735, p < .001$
Weakly-Related Non-Word vs. Unrelated Non-Words	$t(114) = 8.694, p < .001$
Weakly-Related Non-Words vs. Filler(Baseline) Condition	$t(114) = 13.295, p < .001$
Unrelated Non-Words vs. Filler (Baseline) Condition	$t(114) = 4.036, p < .001$

Table 5. Paired-Samples T-Test for Experiment 2

(1)   3 Word Conditions and 3 Non-Word Conditions

Conditions	T-statistic
Strongly-Related Words vs. Strongly-Related Non-Words	$t(111) = -9.694, p < .001$
Weakly-Related Words vs. Weakly-Related Non-Words	$t(111) = 2.868, p = .005$
Unrelated Words vs. Unrelated Non-Words	$t(111) = 2.868, p = .005$
Strongly-Related Words vs. Weakly-Related Words	$t(111) = -6.961, p < .001$
Strongly-Related Words vs. Unrelated Words	$t(111) = -12.548, p < .001$
Weakly-Related Words vs. Weakly-Related Non-Words	$t(111) = -4.895, p < .001$
Strongly-Related Non-Words vs. Weakly Related Non-Words	$t(111) = 1.227, p = .222$
Strongly-Related Non-Words vs. Unrelated Non-Words	$t(111) = 1.451, p = .150$
Weakly Related Non-Word vs. Unrelated Non-Words	$t(111) = .137, p = .891$

(2)   4 Non-Word Conditions

Conditions	T-Statistic
Strongly-Related Non-Words vs. Weakly-Related Non-Words	$t(111) = 1.227, p = .222$
Strongly-Related Non-Words vs. Unrelated Non-Words	$t(111) = 1.451, p = .150$
Strongly-Related Non-Words vs. Filler(Baseline) Conditions	$t(111) = 4.124, p < .001$
Weakly-Related Non-Word vs. Unrelated Non-Words	$t(111) = .137, p = .891$
Weakly-Related Non-Words vs. Filler(Baseline) Condition	$t(111) = 3.342, p = .001$
Unrelated Non-Word vs. Filler (Baseline) Condition	$t(111) = 3.554, p = .001$

The findings did not fully support the third hypothesis, which is that in both experiments, participants would respond more quickly to unrelated words than unrelated non-words. As the above Table 4 and Table 5 demonstrate, the results of a paired samples *t*-test of Experiment 1 did not provide support for this prediction. The results were $t(114)$ = -1.065, p = .289 which did not have any statistical significance although participants made a bit faster response to unrelated words (M = 803.38, SD = 139.48) than unrelated non-words (M = 813.13, SD = 153.94) (Table 1). However the findings from paired samples *t*-test of Experiment 2 seemed to support this prediction, $t(111)$ = 2.868. p = .005, while participants made a bit slower response to unrelated words (M = 810.60, SD = 146.71) than unrelated non-words (M = 785.35, SD = 123.74) (Table 2).

4   Discussion and Conclusion

The present study aimed at evaluating the difference between the response to non-words that look like words and response to non-words that sounds like words. As anticipated, participants make the different responses to non-words that look like words and non-words that sound like words. This finding partially supports the results of Ratcliff, Gomez, & McKoon [2] that participants generally respond more slowly and less accurately to the non-words sharing similar visual characteristics with words. The results support the hypothesis that participants make faster responses to non-words that sound like words than non-words that look like words. For example, the mean RT under the condition of Strongly-Related Non-words of Experiment 1: Looks like Word was 871.58 (SD=140.31) (Table 1) and the mean RT for Experiment 2: Sounds like Word was 759.01(SD=125.21) (Table 2). The visual similarity between words and non-words might have influenced participants' RT to the word strings.

The results also supported the hypothesis that participants responded more slowly to related non-words than related words in both experiments (non-words that look like words and non-words that sound like word). For example, the RT for Strongly-Related Words was 727.80 (SD=105.88) and the RT for Strongly-Related Non-words was 871.58 (SD=140.31) in Experiment 1. As it is discussed in the introduction, Ratcliff, Gomez, & McKoon [2] suggested that drift rate is decided by the familiarity of the presented information. Therefore, the drift rate was significantly higher when the words are familiar for participants. Non-words are less common in daily life. Participants were more likely to have higher familiarity rate toward words rather than non-words and consequently make faster responses in both experiments.

The results did not fully support the hypothesis that participants would make faster responses to unrelated words than unrelated non-words in both experiments. Experiment 1 did not have statistical significance but the participants responded more quickly to unrelated words than to unrelated non-words. By contrast, Experiment 2 had some statistical significance but the participants responded slower to unrelated words than unrelated non-words.

A limitation of the study was that the gender ratio was significantly uneven in Experiment 2, with 32 females and 80 males (after data cleaning). To compare between

the two experiments, the difference between the gender distributions was also significant (after data cleaning, the ratio of females to males was 56:59 in Experiment 1 and 32:80 in Experiment 2, respectively). As a result, differences in gender distributions might have influenced the RT.

Overall, the results of the study suggest that participants tend to make faster responses to non-words that sound like words than non-words that look like words. Moreover, when the words are related strongly, no matter whether it is visually similar or similar in sound, participants tend to respond more slowly to non-words than words. The findings majorly support the study of Ratcliff, Gomez, & McKoon [2] which suggested that visual familiarity can have influential on the results and as a result, participants tend to make faster responses to non-words that look like words. The gender distribution of the study was not controlled evenly, which might have an influence on the study results. Further researches could consider a much more even gender distribution to control the variability between groups. The findings of the study will be beneficial to English lexical teaching and acquisition because they gave transparent evidences on human cognitive processes for lexical differentiation and learning.

5   References

[1] Tillman, G., Osth, A. F., Ravenzwaaij, D., & Heathcote, A. (2007). A diffusion decision model analysis of evidence variability in the lexical decision task. *Psychonomic Bulletin &. Review*, 24.6: 1949-1956. https://doi.org/10.3758/s13423-017-1259-y

[2] Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, 111.1: 159-182. https://doi.org/10.1037/0033295x.111.1.159

[3] Meyer, D. E., & Schvaneveldt, R. W. (1971) . Facilitation in recognizing paires of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90. 2: 227-234. https://doi.org/10.1037/h0031564

[4] Wagenmakers, E.-J., Ratclif, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical. *Journal of Memory and Language*, 58.1: 140-159. https://doi.org/10.1016/j.jml.2007.04.006

[5] Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2018) . Dynamic adjustment of lexical decision task: Cross-trial sequence effects. *Quarterly Journal of Experimental Psychology*, 71.1: 37-45. https://doi.org/10.1080/17470218.2016.1240814

[6] Yap, J. M., Sibley, D., & Rueckl, J. (2014). Responding to Non-words in the lexical decisiontask: insights from the English lexicon roject. *Journal of Experimental Psychology*, 41.3: 597-613. https://doi.org/10.1037/xlm0000064

[7] Lukatela, G., & Turvey, M. T. (1990). Phonemic Similarity Effects and Prelexical Phonology. *Menmory & Cognition,* 18.2: 128-152. https://doi.org/10.3758/bf03197089

[8] Shulman, H. G., Hornak, R., & Sanders, E, (1978) .The effects of graphemic, phonetic, and semantic relationships on access. *Memory & Cognition*, 6.2: 115-123. https://doi.org/10.3758/bf03197436

6   Authors

Shuyu Zhang is an undergraduate student with the Research School of Psychology, Australian National University, Canberra, ACT2600, Australia. u6370961@anu.edu.au

Sihong Zhang is a PhD & Professor in Linguistics, Director of the Centre for Linguistics, School of Foreign Studies, Hefei University of Technology. He works also as a Senior Adjunct Researcher Fellow at the Language & Culture Research Centre, James Cook University, Australia. He has obtained his PhD from James Cook University. His research areas include linguistic typology, language teaching and documentation, and cognitive linguistics.