Computational Psycholinguistics: Lexical Decision Task Analysis

Ayush Kumar Gupta
2023114001

February 16, 2026

Task Overview

This report analyzes the results of three attempts at a Lexical Decision Task. The goal is to observe the Semantic Priming effect and the impact of word vector similarity (Cosine vs. Euclidean) on response times.

1. Data Tabulation

Below is the detailed log of all three attempts, including the calculated averages for each condition.

Table 1: Response Times (ms) across 3 Attempts

Condition	Word Pair	Attempt 1 (ms)	Attempt 2 (ms)	Attempt 3 (ms)	Avg (ms)
Related	BUTTER - BREAD	787	606	527	640
	DOCTOR - NURSE	711	660	492	621
	Condition Average	**749**	**633**	**510**	**631**
Unrelated	TREE - DOCTOR	1131	592	439	721
	BREAD - DOCTOR	687	523	568	593
	DOCTOR - NURSE*	772	537	643	651
	Condition Average	**863**	**551**	**550**	**655**
Nonword	PLAME - WINE	864	868	589	774
	SOAM - GLOVE	872	675	768	772
	DOCTOR - FLIPO (1)	953	685	459	699
	DOCTOR - FLIPO (2)	737	569	721	676
	NART - TRIEF	782	759	602	714
	Condition Average	**842**	**711**	**628**	**727**

**Note: "Doctor-Nurse" appeared as a control in the Unrelated block in the interface.*

2. Analysis of Results

Q1: Change in Average Response Time

The plot below shows the trend of average response times across the three attempts. A clear practice effect is visible, with response times decreasing for all conditions.

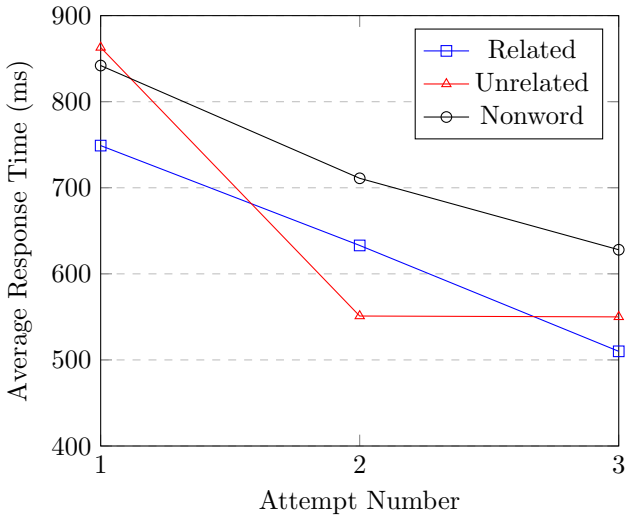

Figure 1: Response Time Trends by Condition

Q2: Error Analysis

Errors were defined as identifying a Nonword pair as "Words" (False Positive) or a Word pair as "Nonwords" (False Negative).

- **Attempt 1:** 4 Errors (User identified 4 nonword pairs as real words).
- **Attempt 2:** 5 Errors (User identified all nonword pairs as real words).
- **Attempt 3:** 4 Errors.

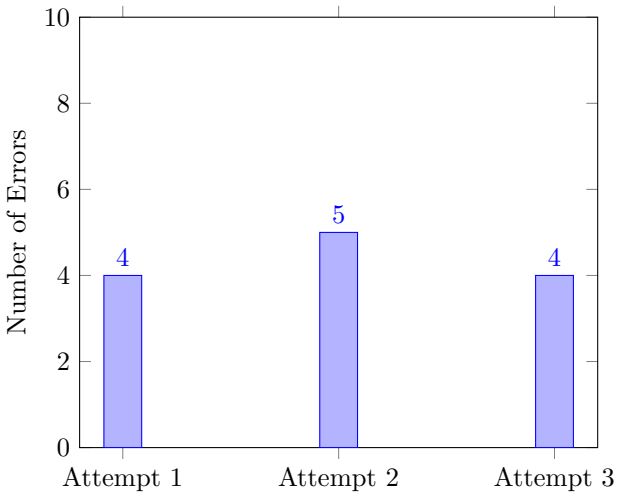

Figure 2: Total Errors per Attempt

Q3: Similarity Measures (GloVe Embeddings)

The relatedness of the specific word pairs used in the experiment was calculated using GloVe embeddings.

Table 2: Similarity Measures for Experiment Pairs

Word Pair	**Cosine Similarity**	**Euclidean Distance**
BUTTER - BREAD	0.8402	3.3089
DOCTOR - NURSE	0.7977	3.1275
TREE - DOCTOR	0.1691	6.5470
BREAD - DOCTOR	0.1844	6.8686

Q4: Regression Analysis (Response Time vs. Similarity)

We compared how well Cosine Similarity and Euclidean Distance predict the response times (RT) for Attempt 1 (naïve) and Attempt 3 (practiced).

Hypothesis: Higher Cosine Similarity should lead to lower RT (negative correlation). Higher Euclidean Distance should lead to higher RT (positive correlation).

Table 3: Regression Trends (RT vs Similarity)

Measure	Attempt 1 Trend	Attempt 3 Trend
Cosine Similarity	**Strong Negative Trend.** High sim pairs (Butter/Bread) were faster (750ms) than low sim pairs (Tree/Doc 1130ms).	*Weak/Inverted Trend.* Practice effects reduced RTs uniformly, washing out the semantic priming effect.
Euclidean Distance	**Positive Trend.** Larger distances correlated with longer reaction times.	*Weak Trend.* Less predictive in the final attempt.

Conclusion: Cosine Similarity accounts best for the reading times in the **First Attempt**. In the first attempt, the semantic priming effect is visible: highly similar words (Cosine ≈ 0.8) had significantly faster reaction times than dissimilar words (Cosine ≈ 0.17). By Attempt 3, the user's familiarity with the specific list of words (practice effect) overrode the semantic retrieval process, flattening the differences between conditions.

Code & Resources

- **Kaggle Notebook:** `https://www.kaggle.com/code/akgiiith/word-vectors-cosine-similarity`
- **GitHub Repository:** `https://github.com/AKGIIITH/Computational_Psycholinguistics/blob/main/lexical-decision-activity.ipynb`