

The Interference of Linguistic Symbols on Categorization

Yuchen Wang*

Boston College, Chestnut Hill, Massachusetts, 02467, United States

* Corresponding Author Email: yuchen.wang.3@bc.edu

Abstract. Categorization, as a fundamental concept that assists people in daily perceiving and estimating, influences the approach to how individuals classify objects, events, and ideas through its language-based presentation. The adaptations of language in science and tradition will inevitably result in paradoxical classifications. To explore to what extent linguistic symbols can interfere in or even determine categorization, this study experimentally tested the ratio of how linguistic interference affects the perceptions of textual, colorific, and melodic information, respectively. Forty-one examinees participated in the experiments, and thirty-two multiple-choice questions were set to quantify the extent of linguistic interference in categorization, as well as trace how linguistic symbols perceived as textual, colorific, and melodic stimuli affect categorization, respectively. The control experiments indicated that linguistic interference, in limited grammatical and verbal forms, could not wholly determine but highly affect categorization, with kappa coefficients of 0.33 for textual, 0.27 for colorific, and 0.24 for melodic ($p < 0.05$). In addition, as shown in the data, linguistic interference had slightly higher effects on the perceptions of untextual information such as colors and melodies.

Keywords: Categorization; Language; Interference.

1. Introduction

Categorization refers to the process that classifies objects, events, and ideas into various taxa based on their features, properties, and functions with the experiences of comparing one to another. Meanwhile, the systematic patterns deriving from the categorizations can govern human lives, even qualifying the material cultures and languages [1]. At the same time, cognitive structures, which can determine the rules to perceive and conceive, are interactively shaped by the existent grammatical and verbal systems of specific languages, and a frame of reference can arise to delineate the boundaries of thoughts and behaviors within specific linguistic communities [2, 3]. Therefore, linguistic meaning is conventional, and categorization through linguistic symbols is restrictive [4]. With the evolution of taxonomic, categorization is extensively applied in different curriculums and industries to organize and navigate discrete items. Investigating how categorization is modified by linguistic symbols can assist people in both conceptual and pragmatic analysis of psychology, sociology, education, etc. The evolvement of categorization from classical views to modern standards is shown in table 1 [5].

Table 1. The Evolution of Categorization

Name	Time	Criteria	Example
Aristotelian's Category	350 BCE	All necessary and sufficient attributes	Bachelor: male, single, and adult
Prototype Theory	1973	Crucial attributes possessed by an idealized prototypical member	Fruit: conforming with crucial attributes of apple, such as terrestrial plants, edible, containing seeds
Exemplar Approach	1978	Resemblance to one or more of the previously defined members	Carp: similar to grass carp, black carp, or koi

Contradictory claims disputing the level of linguistic interference in categorization have been stated since the 1950s. According to the Linguistic Relativity Theory, an extreme version of the theory asserts that languages fully enable and determine the thoughts of speakers. Conversely, a qualified

opinion insists that languages only finitely influence the speakers' minds [6]. Several experiments have been conducted during the past decade to verify how the cognitive processes of categorization are independently modified by textual, colorific, and melodic information. However, the extent to which linguistic symbols interfere in categorization has yet been verified with statistical methods, while the experiments with interactions between linguistic and extralinguistic stimuli can hardly be referred to.

This paper uses control experiments to compare how people react to stimuli with and without linguistic implicatures, and the coherences of the answers, which embody the level of linguistic interference, will be measured by the means, the standard deviations, kappa coefficients, the p-values, and the results of the t-test. In addition, the assessments of the interactions between textual and untextual information are included in the experiments to test the extent to which linguistic symbols impact the perceptions of extralinguistic stimuli.

2. Method

The methods applied in the experiments to verify if linguistic symbols originating through traditional perspectives will interfere in categorization, as well as modify the trajectories of how people define different objects, events, and ideas are interpreted as below.

2.1 Participants

A total of 41 examinees without dyslexia, color blindness, or any perceptual disorders will participate in this study via the online platform PsychoPy/PsychoJS, while all the participants are college students who are fluent in English and range in different genders between the ages of 19 to 25.

2.2 Procedures

The experiments will test the degree of linguistic interference in categorization by comparing the responses of participants with or without linguistic guides. Specifically, the means, the standard deviations, kappa coefficients, the p-values, and the results of the t-test will be computed to reveal the extent to which linguistic symbols can modify the perceptions of both textual and untextual information through visual and auditive stimuli. The formulas that interpret the significant algorithms computing the co-selection rates between the answers with and without linguistic interferences are shown below [7, 8].

For the kappa coefficient, the value k refers to the kappa coefficient, while p_0 is the co-selection rate and p_e is the hypothetical probability of the options (0.25 for each of the four options).

$$k = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \quad (1)$$

For the t-test, the value t refers to the t-value, while \bar{x}_1 and \bar{x}_2 are respectively the means of the two groups of answers, and s^2 is the squared standard of these two groups. Finally, n_1 and n_2 are the numbers of participants in the experiments.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]}} \quad (2)$$

For the p-value, it equals the product of the probability p and the arithmetic mean of the test statistic TS calculated from the samples and the null hypothesis H_0 .

$$p - value = p(TS|H_0) \quad (3)$$

As the types of questions vary in each section, the specific rules to score the answers will be explained in detail in the section on experimental results below.

2.2.1 Task One

In the first section, participants will initially answer the multiple-choice questions with the option combinations represented in distorted linguistic symbols, and four options in the combinations consist of a correct answer, an interferential answer, and two incorrect answers. Then, the same questions will appear with the options that include the correct answer, the interferential answer, and the other two incorrect answers in neutral lexical structures without implications. These typical questions contain obscure knowledge which can be easily misled by languages, and the responses will be compared to track the interferences caused by linguistic symbols to categorization. The questions and two sets of options are listed in table 2, interferential options are underlined, and optimal options are marked in bold.

Table 2. Textual Section of the Experiment

	<i>Question</i>	<i>Option Combination 1</i>	<i>Option Combination 2</i>
<i>Set 1st</i>	Which marine organism below is not a cephalopod (a mollusk with a prominent head and a set of tentacles)?	A. Jellyfish B. Ink-fish C. <u>Octopus</u> D. Cuttlefish	A. <u>Octopus</u> B. Squid C. Jellyfish D. Nautilus
<i>Set 2nd</i>	Which below is not a Scandinavian country?	A. Finland B. Netherlands C. <u>Denmark</u> D. Iceland	A. Norway B. Sweden C. <u>Denmark</u> D. Netherlands
<i>Set 3rd</i>	Which super-hero below does not have any extraordinary powers?	A. Spiderman B. Batman C. Antman D. <u>Captain Marvel</u>	A. <u>Captain Marvel</u> B. Doctor Strange C. Batman D. Hulk
<i>Set 4th</i>	Which number below do you think attributively differs from the others?	A. <u>Twelve</u> B. Fourteen C. Sixteen D. Seventeen	A. <u>12</u> B. 14 C. 16 D. 17

2.2.2 Task Two

The second section will test how linguistic symbols, which possess representational meanings, will affect the categorization of the sensory stimuli generated from extralinguistic information (colorific and melodic) [9, 10].

First, eight gradient color blocks between blue and green will be successively manifested with the true names (shown on the left), and participants will spontaneously select whether each block is blue or green with the key “b” or “g”. Afterward, the blocks will randomly appear with the false names (shown on the right), and participants will make the choices again. Between the two stages of the task, the names of the blocks in similar shades of colors are interchanged. Precisely, Spring Mint and Adriatic Mist, Tide Water and Eucalyptus Leaf, Seafoam Pearl and Botanical Tint, and Verdigris and Capri Sea are four pairs to switch the names of color blocks. The gradient colors with the true and false names are illustrated in figure 1.



Figure 1. Color blocks with true and false names.

Furthermore, four audio clips that consist of vocal, instrumental, or synthetic components are displayed to test if the vocal parts with textual lyrics dominate over the instrumental parts without any textual information. The vocal and instrumental elements are respectively snipped from different musical compositions but with congruent melodies. Coincidentally, all the vocal parts are clipped from pop music, while all the instrumental parts are clipped from classical music. As the sequences to play the vocal and instrumental clips in each set may cause judgmental selections, two experimental sets will only play the instrumental clips before the synthetic clips, while the vocal clips will only be played before synthetic clips in another two experimental sets. Finally, participants will be inquired whether each clip sounds like classical or pop, and the answers can be submitted with the key “c” or “p”. Detailed components of each audio clip are listed in table 3.

Table 3. Audio Clips.

	<i>Vocal</i>	<i>Instrumental</i>
<i>Set 1st</i>	N/A	Johann Pachelbel’s Canon in D
	Maroon 5’s Memory	Johann Pachelbel’s Canon in D
<i>Set 2nd</i>	N/A	Ludwig van Beethoven’s Moonlight Sonata
	The Beatles’ Because	Ludwig van Beethoven’s Moonlight Sonata
<i>Set 3rd</i>	Lady Gaga’s Alejandro	N/A
	Lady Gaga’s Alejandro	Vittorio Monti’s Czardas
<i>Set 4th</i>	Celine Dion’s All by Myself	N/A
	Sergei Rachmaninoff’s Piano Concerto No. 2	Sergei Rachmaninoff’s Piano Concerto No. 2

3. Experimental Results

This section will reflect the degrees of interferences caused by linguistic symbols in textual and multiple sensory forms (with visual and auditive). Based on the performances of participants, different algorithms with the means, the standard deviations, the kappa coefficients, the results of the t-test, and the p-values will be calculated to contrast how participants differently respond to the questions with and without linguistic interference.

3.1 The Experimental Result of Section One

In the first section, the selection rates of the correct and interferential options in option combinations one and two are computed. Specifically, this section consists of eight multiple-choice questions with four options each, and the correct, interferential, and other incorrect options are assigned various scores to compare the consistencies of how participants respond to the same questions with disparate combinations of options. The kappa coefficient of 0.33 (N=41) illustrates that the participants choose different answers from the disparate combinations of options for the same questions, while the results of the t-test demonstrate increased accuracies without linguistic interference. For the correct options, a significant discrepancy is shown with a kappa coefficient of 0.33. Moreover, with a p-value less than 0.05, the independent sample t-tests, which indicate the differences between the collected scores of the two combinations, are respectively computed as 7.30 and 8.76. On the contrary, the selection rates of interferential options descend from option combination one to two. Precisely, the kappa coefficient equals 0.36 (N=41), and the value of independent sample t-tests for option combinations one and two decreases from 6.22 to 4.42 with a p-value less than 0.05. A histogram compares all the values above is shown below in Figure 2.

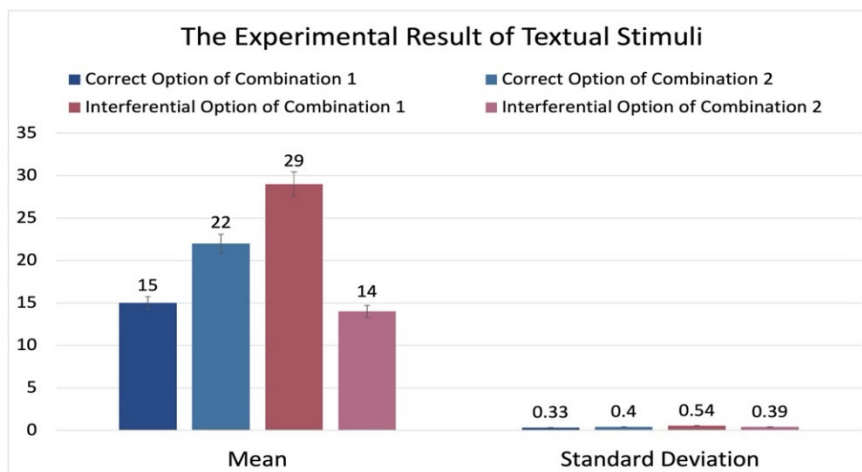


Figure 2. The experimental result of section one.

3.2 The Experimental Result of Section Two

The performances in section two reveal a more distinct discrepancy that indicates the linguistic interference to colorific and melodic information with kappa coefficients of 0.27 and 0.24, respectively. Additionally, with a p-value less than 0.05, the t-tests referring to the overlapping ratios of the answers are successively calculated as 6.82 and 8.56 for colorific stimuli and 3.62 and 9.33 for melodic stimuli. Specifically, because all the questions have binary choices in the section, the answers with and without linguistic interference are directly compared. As the colorific and melodic stimuli are untextual, this section can verify how linguistic symbols interfere in categorizing untextual information. The results of kappa coefficients demonstrate that the participants make different choices following linguistic implications, and the increased values of t-tests show that linguistic symbols can highly affect people’s decisions on the categorization of untextual information. Two histograms that compare all the values above are shown below in Figure 3.

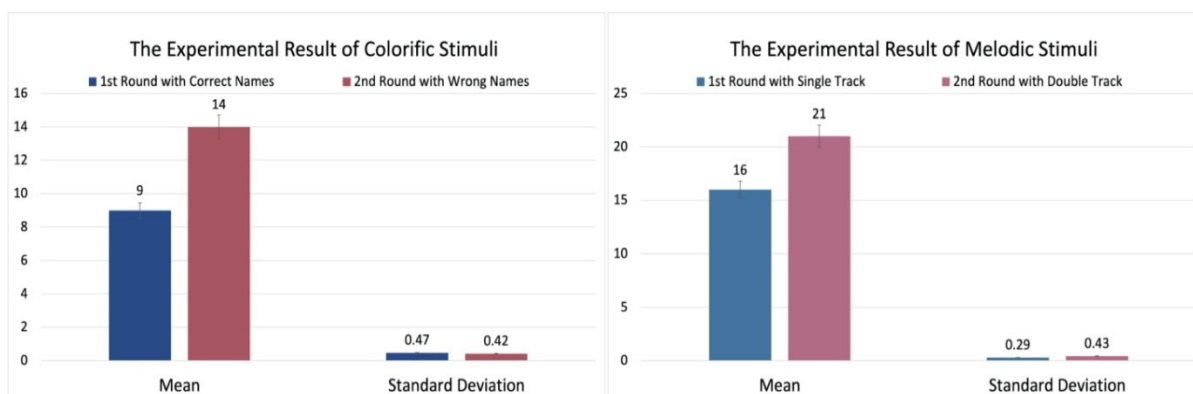


Figure 3. The experimental result of section two.

4. Discussion

The experiments provide a quantitative representation of the controversial hypothesis about the extent to which language interferes with categorization, as well as demonstrate how symbolic languages shape the ways people percept sensory stimuli formed by untextual information such as colors and melodies. For the textual multiple-choice questions, various scores are weighted to correct, interferential, and other incorrect options. The extent to which linguistic interference in categorization is demonstrated by comparing participants’ scores to the same questions with different combinations of options. For task one with textual information, the experimental results show that the correct answers are chosen more frequently from the second option combinations without misleading

linguistic symbols than the first option combinations under linguistic interference, while the inferential options appear less frequently in the neutral option combinations two than the distorted option combinations one. Therefore, slightly higher scores are acquired from the second option combinations without linguistic interference than the first option combinations interfered with linguistic symbols. In the second task of binary-choice questions, the participants' performances are affected by the presence or absence of linguistic interference. The experimental results indicate that the names with inclined hints can induce people to predetermine the colors of the blocks when colorific stimuli cannot lead to confirmatory judgments. Furthermore, the musical clips with almost identical melodies are distinguished into binary genres as the linguistic stimuli in the vocal tracks whelm the extralinguistic stimuli in the instrumental tracks. The experimental results verify the qualified version of the assumption stated in the Whorf hypothesis, which means the linguistic interference is prominent but still restrictive. Within the limited scales of the experimental samples and questions, the radical version of the assumption can be denied since the perceptions of information in multiple formats cannot be entirely determined by languages. In other words, the procedures of how people perceive and conceive the information may successively occur in the different intervals of their brains [11].

However, like plenty of other studies, several limitations exist in this experiment. First, the generalization is limited by the small amounts of representative questions, as the complexity of the grammatical and verbal interferences cannot be completely embodied by the samples. How to mathematically differ the scores of the correct, inferential, and other incorrect answers in the multiple-choice questions for data analysis is another challenge. Typically, more statistical methods can be applied if the size of samples is increased. In future experiments, a larger size of samples containing the interferences from different linguistic elements should be set necessarily. Furthermore, subsequent experiments that attempt to analyze the specific steps in processing linguistic information and categorizing can be created to explore whether memory retrieval participates in the procedures and which areas of the human brain the cognitive activities are accomplished [12].

5. Conclusion

As mentioned above, the linguistic interference in categorization, as a consensus of cognitive science, lacks the quantitative research to verify the level of interferences caused by linguistic symbols, and the controversy between whether linguistic interferences determine or finitely affect categorization has yet been resolved. This experiment aims to use quantitative methods to calculate and analyze the interferences originating from some representative linguistic conventions with the samples on a limited scale. In addition, further questions may be raised, such as how the deviations caused by linguistic interference in categorization will affect the structures of other disciplines, while the processes of how people categorize with and without linguistic interference deserve to be technically explored in the future. Long-term memory retrieval may be a significant factor influencing the relevance between the perception of linguistic information and categorization since the procedures to classify objects, events, and ideas are triggered by long-term memory retrievals that can match the external sensory stimuli and engrams generated and stored internally. The experiments that trace and compare the different approaches to categorizing with and without linguistic interference may be set in the future. Moreover, contrast experiments may test the discrepancies in categorization caused by different language systems.

References

- [1] Brown, R. W., & Lenneberg, E. H. A study in language and cognition. *The Journal of Abnormal and Social Psychology*, 1954, 49(3), 454.
- [2] Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. Can language restructure cognition? *Trends in cognitive sciences*, 2004, 8(3), 108-114.

- [3] Fountain, T., & Lapata, M. Meaning representation in natural language categorization. In Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32), 2010.
- [4] Narens, L., Jameson, K. A., Komarova, N. L., & Tauber, S. Language, categorization, and convention. *Advances in Complex Systems*, 15(3), 2012.
- [5] Smith, E. E., Patalano, A. L., & Jonides, J. Alternative strategies of categorization. *Cognition*, 1998, 65(2-3), 167-196.
- [6] Whorf, B. L. (2012). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. MIT Press.
- [7] Uebersax, J. S. A Generalized kappa coefficient. *Educational and Psychological Measurement*, 1982, 42(1), 181-183.
- [8] Doug, S. Tests and measurements: the t-test. *Strength and Conditioning Journal*, 1990, 12(1), 36-37.
- [9] Ozgen, E. Language, learning, and color perception. *Current Directions in Psychological Science*, 2004, 13(3), 95-98.
- [10] Slevc, R. L. Language and music: sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2012, 3(4), 483-492.
- [11] Schoot, L., Hagoort, P., & Segaert, K. What can we learn from a two-brain approach to verbal interaction? *Neuroscience & Biobehavioral Reviews*, 2016, 68, 454-459.
- [12] Juslin, P., Jones, S., Olsson, H., & Winman, A. Cue abstraction and exemplar memory in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2003, 29(5), 924.