

The role of word frequency and contextual diversity in visual word recognition: a mini review

Xin Huang*

School of Psychology, Nanjing Normal University, Nanjing, China

Abstract

Contextual diversity refers to the number of contexts in which a word appears. It is traditionally believed that word frequency is an important factor affecting lexical access, but the presence of contextual diversity challenges the position of word frequency in lexical cognition. The research on contextual diversity and word frequency is mainly focused on two fields, namely word recognition and word learning. However, research outcomes concerning the mapping between contextual diversity and word frequency are inconsistent with each other: 1) Contextual diversity can replace word frequency, and is better than word frequency in terms of word recognition; 2) Word frequency and contextual diversity are two different variables that affect word recognition, both of which independently affect word recognition. No definite conclusion about the relationship between the two has been reached yet. Finally, we highlight topics that are in need of future systematic research.

Introduction

The effect of word frequency on lexical processes is both ubiquitous and large. High-frequency words are known to more people and are processed faster than low-frequency words [1-4]. Also, many evidence from behavior, electrophysiological and neural image data all proved the existence of word frequency [5-8]. Besides the evidence from corpus of Indo-European languages, word frequency is also found in the corpus of Sino-Tibetan languages [9]. This is the case both in adult readers [10,11] and in young readers [12,13].

Word frequency typically explains some 30-40% of the variance in almost visual word and reading models. For example, in the family of localist activation-based models, the resting level of activation of a given word unit depends on its printed word frequency [7,14-16]. Likewise, computational models of reading such as the E-Z Reader model [17] and the SWIFT model [18] employ a similar mechanism of word frequency during the initial "lexical access" stage.

However, with the deepening of the study of word frequency effect, more and more scholars have doubts about the way of action of word frequency: frequency is core to classic strength accounts of lexical access based on the assumption that each repetition increases memory strength for a word, boosting the efficiency of later access. However recent research on pure repetition (repetitive effects do not depend on other factors) showed that repetitive effects were always mixed with other factors, like time [19] and contexts [20] and thus influenced word recognition. It means that word frequency may affect lexical process with other factors and thus challenged the role of word frequency in word recognition. In sum, research has proved that word frequency plays an important role in word recognition, while with the deepening of research, the way of word frequency effect is still unclear, so contextual diversity came into interpreting lexical processing.

Adelman, Brown and Quesada operationalized a measure called contextual diversity (CD) [21], it refers the number of contexts in a corpus in which a word is experienced. Brysbaert and New B [22] then gave an operational definition, it refers to the number of movies or

dramas that a word appears in a subtitle corpus, and further research almost use subtitles as source of corpus, and it has been proved to explain significantly more of the variance in word naming and lexical decision performance than the measures based on written texts [23-26].

The concept of contextual diversity is proposed by many researchers, while it also raised controversy. Johns, Gruenenfelder, Pisoni and Jones [27] questioned that the definition of contextual diversity ignores the information redundancy of words in contexts. Besides, it is questionable whether common operational definitions of CD are valid measures under the principle of likely need. Furthermore, context has been described in different ways [28]. For example, context has been described as information that fluctuates randomly over time with respect to different item presentations [29-31] which might be referred to as temporal context [32]. Context has also been defined as the physical environment in which an item occurs [33-35]. Howard and Kahana (2002) described the context associated with a given item as a composite representation of the semantic features of the items that preceded it on a list [36]. It is not directly obvious how counting documents corresponds to classic notions of a change in context, but it seems intuitive that if a document is repeated in the corpus, we should not consider the two repetitions to be different semantic contexts of the word.

Based on the discussion above, the relationship of word frequency and contextual diversity in visual word recognition is still unclear, and no strong evidence showed the way they influenced lexical processing. We begin our review by discussing the recent research on word frequency and contextual diversity in visual word recognition, including methods and relevant research. We conclude with issues that call for future systematic research.

Correspondence to: Xin Huang, School of Psychology, Nanjing Normal University, Nanjing, China, 210097. E-mail: huangxincoral@foxmail.com

Key words: contextual diversity, word frequency, word recognition, word learning

Received: November 07, 2017; **Accepted:** November 27, 2017; **Published:** November 30, 2017

Methods

Contextual diversity and word frequency in word recognition

Most researchers carried out their research by lexical decision task [1,37] and word naming task to explore the role of word frequency and contextual diversity in visual word recognition. Lexical decision task is the most common laboratory task for studying word recognition. Participants are required to decide whether a string of letters is a word or not (a nonword) [38-40].

Word naming task is another common laboratory task. This method requires participants to read out a letter or a word, or name a word, and record the naming latency. Word naming can measure the time required to identify a word, as well as various factors that affect word recognition. The frequency of the word, the degree of complexity, whether the background context and other factors have a great impact on the naming time. Compared to vocabulary lexical decision tasks, word naming usually occurs at the later stage of lexical access. In addition, it is worth noting that it may appear the separation of speech and semantic phenomenon when people name a word, such as patients with traumatic brain injury can name and distinguish between words and non-words, but can not understand the meaning of words, thus word naming tasks cannot guarantee lexical access. Therefore, the use of naming tasks is less than the vocabulary judgment task in visual word recognition.

Relevant research on word frequency and contextual diversity

Adelman [1] first investigate the role of word frequency and contextual diversity by lexical decision and word naming tasks. He found, when contextual diversity is control, word frequency effect disappears; when controlling the frequency of words, the contextual diversity effect still exists, so he believes that the context of diversity is a better predictor than word frequency in word recognition. The developmental research by Perea *et al.* [41] also supported Adelman's view: he selected 22 fourth-year children in Portuguese, and likewise used lexical decision task to get the similar result and thus extending the contextual effects to Fourth grade children. In the sentence reading, there are similar findings, for example, Plummer, Perea and Rayner [41] used eye movement to find that, all low CD words reflect the early processing of eye movement indicators such as the first fixation time, single gaze time after controlling the word frequency, and the eye movement indicators, such as the retrospective path time and the total gaze time, were significantly higher than those of the high frequency words. After controlling the diversity of the situation, the high and low frequency words had no significant difference in all eye movements. Chen *et al.* [42] found the same results after they get refinement on Plummer's research. The studies above show that contextual diversity is a better predictor than word frequency in word recognition.

However, Vergara-Martínez *et al.*, [26] found, when using ERPs to detect the role of contextual diversity, it is not an incidental phenomenon of word frequency, and both cannot be substituted for visual word recognition in reading. This contradictory result suggests that the mechanism of contextual diversity is still unclear, and existing research cannot provide clear and strong scientific evidence. First, the Vergara-Martínez *et al.* [26] did not control the lexical semantic diversity, making the driving mechanism of the contextual diversity effect unclear; secondly, although the EEG technique had a very high temporal sensitivity, Vergara-Martínez *et al.* did not control the materials (words high-CD high WF and words with low CD high WF word are marginal significant in the frequency of words, $p = 0.07$, Figure. 1), so that it cannot be concluded that the observed effects of

	CD comparison			
	Range	High CD	← p →	Low CD
Contextual diversity ^a	0–100	41.8 (4.6)	5.3 ⁻¹⁶	27.7 (4.4)
Subtitle-frequency ^b	0–33,170	155.1 (5.2)	.07	139.3 (5.1)
imageability	1.2–7	4.9	.12	5.4
Concreteness	1.9–6.8	4.7	.15	5.1
Age of acquisition ^c	1–11	4.6	.14	5.2
Number of letters	4–8	5.5	.52	5.7
Levenshtein distance	1–12	1.5	.30	1.6
Mean positional bigram frequency (token)	0–58,827	6584	.29	5535
Most frequent lexical category		Noun = 29		Noun = 29
		Verb = 5		Verb = 5
		Adjective = 1		Adjective = 1

Figure 1. Mean values for sublexical, lexical, and semantic characteristics of the stimuli [26]

contextual diversity have no interference from word frequency. Given the contradictory results: (1) Contextual diversity rather than word frequency is a better predictor of word recognition [1, 41-43]. (2) both of them could influence word recognition independently, and they are two total different variables [26, 44], thus need more research and discussion to confirm the relationship of the two.

Conclusions and future directions

As described above, research on the relationship between contextual diversity and word frequency is mainly focused on word recognition, but the contradictory results of the effects of the two on the word recognition process showed that the relationship between the two is still not very clear. Basically speaking, Word frequency is highly correlated with context frequency ($R = 0.98$)

Relationship between word frequency and context frequency (both shown on logarithmic scale). Gray points show relationship for all words in the Touchstone Applied Science Associates Corpus (Figure. 2)[45].

However, there are not enough words with low WF and high CD, thus making existing research can only discard this condition. The incompleteness of experimental material may lead to the incompleteness of the comparison result. Therefore, one of the important means to explore the relationship between the word frequency and contextual diversity is to make the two match in different levels. Besides, White *et al.* [46] found that word frequency only showed linear effect in skipping a word and on first fixation duration, while no linear effect in other measures. Furthermore, the effect of contextual diversity has never before been induced experimentally; to do so would require control over the statistical structure of the language being learned. Finally, the existing research on word frequency and contextual diversity mainly divides the two into high and low levels, rather than categorical or a continuous predictor, so how the two variables affect word recognition still need further support. Therefore, we point to several areas in which more research is needed.

First, we may understand the role of word frequency and contextual diversity in word recognition from other domains. The existing research on word learning may provide a new vision. Contextual diversity has been claimed to be a relevant factor to word acquisition in developing readers [47]. Hills *et al.*, [47] examined the co-occurrence of words in caregiver speech from the CHILDES database and found a word's contextual diversity predicted the order of early word learning and

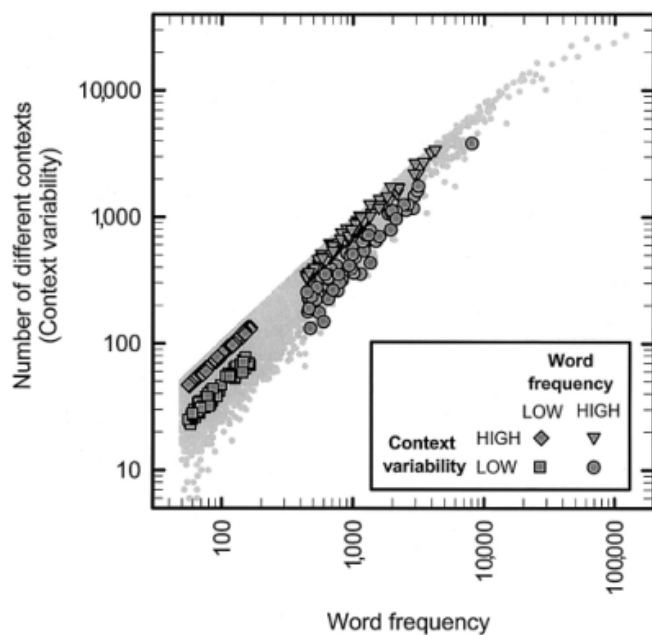


Figure 2. Relationship between word frequency and context frequency (both shown on logarithmic scale). Gray points show relationship for all words in the Touchstone Applied Science Associates Corpus [45].

was highly correlated with the number of unique associative cues for a given target word in adult free association norms. Johns *et al.* [48] using natural language learning paradigm found that when novel words were encountered across distinct discourse contexts, subjects were both faster and more accurate at recognizing them than when they were seen in redundant contexts. However, learning across redundant contexts promoted the development of more stable semantic representations.

Furthermore, memory studies related to contextual diversity can also provide another new perspective. Parmentier, Comesaña and Soares [49] found a total different result with Adelman's research in serial recall performance, that is, the effect of word frequency and contextual diversity disentangled. To be specific, when contextual diversity is controlled, the word frequency effect is still present in the serial recall task, and the two effects are independent in memory performance, while the performance of words with high frequency and low contextual diversity is best. It is not clear that if we can get similar effects in other recall tasks. So the comparison of the two effects in more field and experimental paradigm may a better perspective and approach provide to clarify the relationship between the two.

Secondly, given the reality that word frequency is highly correlated with a number of other word features: word length, age at which the word was acquired, similarity to other words, other factors underlying word frequency and contextual diversity may influence word recognition. For example, Johns *et al.* [50] brought the concept of semantic diversity, it refers the account the number of different semantic contexts in which the word appears [48,50,51].

However, there are still objections on semantic diversity. First, semantic diversity is not much different from semantic richness used in existing research. Semantic richness is a multidimensional structure, including the number of semantic neighbors (NSN) of a word and the number of features associated with the indicator and its contextual dispersion (CD). Second, Plummer *et al.* [41] found that after controlling word frequency and semantic diversity, all the first

pass words and subsequent reading times in words with low CD were higher than those of high CD words. After controlling the contextual diversity, there was no significant difference in the reading time of the low frequency words. This study strongly suggests that the influence of contextual diversity and word frequency on reading and word decision is not affected by semantic diversity.

Furthermore, if there are other factors that may influence word frequency and contextual diversity together, that is, if there are some underlying factors under the two variables. The meager evidence existed cannot proved it.

Finally, EEG, fMRI and other related electrophysiological means can provide us with important indicators for contextual diversity and word frequency in terms of word recognition. And research on patients with traumatic brain injury and other groups, may provide us new brain physiological mechanism perspective to reveal the relationship between the two in the word recognition. Previous studies have shown that word frequency effects appear in the memory of the mirror effect [52], that is, classes of stimuli that are accurately recognized as old when old are also accurately recognized as new when new; those that are poorly recognized as old when old are also poorly recognized as new when new. This phenomenon occurs in patients with Alzheimer's dementia [53,54], Korakov amnesia [55], midazolam amnesia [56] and other cognitive impairment in patients are widespread. In addition, the identification of word frequency effects in patients with schizophrenic patients such as schizophrenic patients is found to be more difficult to identify high frequency words than in low frequency words, possibly because high frequency words may be stored in long-term memory [57]. However, there has been no further study of the relationship between word frequency and contextual diversity from physiological mechanism perspective, which may limit our understanding of word recognition.

In conclusion, the relationship between contextual diversity and word frequency in the field of word recognition still needs further proof and explanation. Whether they could be replaced or have independent influence still needs further research to prove.

References

- Murnane K, Phelps MP, Malmberg K (1999) Context-dependent recognition memory: the ICE theory. *J Exp Psychol Gen* 128: 403-415. [Crossref]
- Balota DA, Cortese MJ, Sergent-Marshall SD, Spieler DH, Yap M (2004) Visual word recognition of single-syllable words. *J Exp Psychol Gen* 133: 283-316. [Crossref]
- Garlock VM, Walley AC, Metsala, JL (2001) Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and language* 45: 468-492.
- Frederiksen JR, Kroll JF (1976) Spelling and sound: approaches to the internal lexicon. *Journal of Experimental Psychology Human Perception & Performance* 2: 361-379.
- Fiebach KI, Chambers SM (1973) Lexical access and naming time 1. *Journal of Verbal Learning & Verbal Behavior* 12: 627-635.
- Reichle ED, Pollatsek A, Fisher DL, Rayner K (1998) Toward a model of eye movement control in reading. *Psychol Rev* 105: 125-157. [Crossref]
- Rugg MD, Doyle MC (1992) Event-related potentials and recognition memory for low- and high-frequency words. *J Cogn Neurosci* 4: 69-79. [Crossref]
- Shiffrin RM, Steyvers M (1997) A model for recognition memory: REM-retrieving effectively from memory. *Psychon Bull Rev* 4: 145-166. [Crossref]
- Rumelhart DE, McClelland JL (1981) An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychol Rev* 88: 375-407. [Crossref]
- Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, et al. (2007) The English Lexicon Project. *Behav Res Methods* 39: 445-459. [Crossref]

11. Johns BT, Dye M2, Jones MN2 (2016) The influence of contextual diversity on word learning. *Psychon Bull Rev* 23: 1214-1220. [[Crossref](#)]
12. Inhoff AW, Rayner K (1986) Parafoveal word processing during eye fixations in reading: effects of word frequency. *Percept Psychophys* 40: 431-439. [[Crossref](#)]
13. Chen Q, Huang X, Bai L, Xu X, Yang Y, et al. (2016) The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychon Bull Rev* 1-9. [[Crossref](#)]
14. Dennis S, Humphreys MS (2001) A context noise model of episodic word recognition. *Psychol Rev* 108: 452-478. [[Crossref](#)]
15. Soares AP, Machado J, Costa A, Iriarte Á, Simões A, et al. (2015) On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *Q J Exp Psychol (Hove)* 68: 680-696. [[Crossref](#)]
16. Megeoch, JA (1942) The psychology of human learning: an introduction.
17. Scarborough DL, Cortese C, Scarborough HS (1977) Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology Human Perception & Performance* 3: 1-17.
18. Fiebach CJ, Friederici AD, Müller K, von Cramon DY (2002) fMRI evidence for dual routes to the mental lexicon in visual word recognition. *J Cogn Neurosci* 14: 11-23. [[Crossref](#)]
19. Smith SM, Glenberg A, Bjork RA (1978) Environmental context and human memory. *Memory & Cognition* 6: 342-353.
20. Duchon A, Perea M, Sebastián-Gallés N, Martí A, Carreiras M (2013) EsPal: one-stop shopping for Spanish word properties. *Behav Res Methods* 45: 1246-1258. [[Crossref](#)]
21. Adelman JS, Brown GD, Quesada JF (2006) Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol Sci* 17: 814-823. [[Crossref](#)]
22. Cai Q, Brysbaert M (2010) SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One* 5: e10729. [[Crossref](#)]
23. Carmen Moret-Tatay, Manuel Perea (2011) Do serifs provide an advantage in the recognition of written words? *Journal of Cognitive Psychology* 23: 619-624.
24. Daltrozzo J, Schön D (2009) Conceptual processing in music as revealed by N400 effects on words and musical targets. *J Cogn Neurosci* 21: 1882-1892. [[Crossref](#)]
25. Engbert R., Nuthmann A, Richter EM, Kliegl R (2005) Swift: a dynamical model of saccade generation during reading. *Psychol Rev* 112: 777-813. [[Crossref](#)]
26. White SJ, Drieghe D, Liversedge SP, Staub A (2016) The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *Q J Exp Psychol (Hove)*. [[Crossref](#)]
27. Johns BT, Dye M, Jones MN (2014) The influence of contextual variability on word learning. In P. Bello, M. Guarani, M. McShane, & B. Scassellati (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 242–247). Austin: Cognitive Science Society.
28. Mensink GJ, Raaijmakers, JG (1988) A model for interference and forgetting. *Psychological Review* 95: 434-455.
29. Brysbaert M, New B (2009) Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods* 41: 977-990. [[Crossref](#)]
30. Monsell S, Doyle MC, Haggard PN (1989) Effects of frequency on visual word recognition tasks: where are they? *J Exp Psychol Gen* 118: 43-71. [[Crossref](#)]
31. Steyvers M, Malmberg KJ (2003) The effect of normative context variability on recognition memory. *J Exp Psychol Learn Mem Cogn* 29: 760. [[Crossref](#)]
32. Hirshman E, Fisher J, Henthorn T, Arndt J, Passannante A (2002) Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language* 47: 499-516.
33. Hofmann MJ, Herrmann MJ, Dan I, Obring H, Conrad M, et al. (2008) Differential activation of frontal and parietal regions during visual word recognition: an optical topography study. *Neuroimage* 40: 1340-1349. [[Crossref](#)]
34. Parmentier FB, Comesaña M, Soares AP (2016) Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Q J Exp Psychol (Hove)* [[Crossref](#)]
35. Steyvers M, Malmberg KJ (2003) The effect of normative context variability on recognition memory. *J Exp Psychol Learn Mem Cogn* 29: 760. [[Crossref](#)]
36. Hyönä J, Olson RK (1995) Eye fixation patterns among dyslexic and normal readers: effects of word length and word frequency. *J Exp Psychol Learn Mem Cogn* 21: 1430. [[Crossref](#)]
37. Chen Q, Huang X, Bai L, Xu X, Yang Y, et al. (2016) The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychon Bull Rev* 1-9. [[Crossref](#)]
38. Davis CJ (2010) The spatial coding model of visual word identification. *Psychol Rev* 117: 713-758. [[Crossref](#)]
39. Glanzer M, Adams JK (1990) The mirror effect in recognition memory: data and theory. *J Exp Psychol Learn Mem Cogn* 16: 5-16. [[Crossref](#)]
40. Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J (2001) DRC: A dual route cascade model of visual word recognition and reading aloud. *Psychol Rev* 108: 204–256 [[Crossref](#)]
41. Polich J, Donchin E (1988) P300 and the word frequency effect. *Electroencephalogr Clin Neurophysiol* 70: 33-45. [[Crossref](#)]
42. Cleland AA, Gaskell MG, Quinlan PT, Tamminen, J (2006) Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *J Exp Psychol Hum Percept Perform* 32: 104. [[Crossref](#)]
43. Hills TT, Maouene J, Riordan B, Smith LB (2010) The Associative Structure of Language: Contextual Diversity in Early Word Learning. *J Mem Lang* 63: 259-273. [[Crossref](#)]
44. Johns BT, Gruenenfelder TM, Pisoni DB, Jones MN (2012) Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *J Acoust Soc Am* 132: EL74-80. [[Crossref](#)]
45. Van Heuven WJ, Mandera P, Keuleers E, Brysbaert M (2014) SUBTLEX-UK: a new and improved word frequency database for British English. *Q J Exp Psychol (Hove)* 67: 1176-1190. [[Crossref](#)]
46. Wilson RS, Bacon LD, Fox JH, Kramer RL, Kaszniak AW (1983) Word frequency effect and recognition memory in dementia of the Alzheimer type. *J Clin Neuropsychol* 5: 97-104. [[Crossref](#)]
47. Hintzman DL, Block RA, Summers JJ (1973) Contextual associations and memory for serial position. *Journal of Experimental Psychology* 97: 220-229.
48. Jones MN, Johns BT, Recchia G (2012) The role of semantic diversity in lexical organization. *Can J Exp Psychol* 66: 115. [[Crossref](#)]
49. Perea M, Soares AP, Comesaña M (2013) Contextual diversity is a main determinant of word identification times in young readers. *J Exp Child Psychol* 116: 37-44. [[Crossref](#)]
50. Johns BT, Dye M, Jones MN (2014) The influence of contextual variability on word learning. In P. Bello, M. Guarani, M. McShane, & B. Scassellati (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 242–247). Austin: Cognitive Science Society.
51. Johns BT, Gruenenfelder TM, Pisoni DB, Jones MN (2012) Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *J Acoust Soc Am* 132: EL74-80. [[Crossref](#)]
52. Godden DR, Baddeley AD (1975) Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology* 66: 325-331.
53. Balota DA, Burgess GC, Cortese MJ, Adams DR (2002) The word-frequency mirror effect in young, old, and early-stage alzheimer's disease: evidence for two processes in episodic recognition performance. *Journal of Memory & Language* 46: 199-226.
54. White SJ, Drieghe D, Liversedge SP, Staub A (2016) The word frequency effect during sentence reading: A linear or nonlinear effect of log frequency? *Q J Exp Psychol (Hove)*. [[Crossref](#)]
55. Plummer P, Perea, M, Rayner K (2014) The influence of contextual diversity on eye movements in reading. *J Exp Psychol Learn Mem Cogn* 40: 275-283. [[Crossref](#)]
56. Howard MW, Kahana MJ (2002) A distributed representation of temporal context. *Journal of Mathematical Psychology* 46: 269-299.
57. Brébion G, David AS, Bressan RA, Pilowsky LS (2005) Word frequency effects on free recall and recognition in patients with schizophrenia. *J Psychiatr Res* 39: 215-222. [[Crossref](#)]

Copyright: ©2017 Huang X. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.