Journal of Eye Movement Research
3(3):3, 1-10

Object Frequency and Predictability Effects on Eye Fixation Durations in Real-World Scene Viewing

Hsueh-Cheng Wang Alex D. Hwang Marc Pomplun
University of Massachusetts at Boston

During text reading, the durations of eye fixations decrease with greater frequency and predictability of the currently fixated word (Rayner, 1998; 2009). However, it has not been tested whether those results also apply to scene viewing. We computed object frequency and predictability from both linguistic and visual scene analysis (LabelMe, Russell et al., 2008), and Latent Semantic Analysis (Landauer et al., 1998) was applied to estimate predictability. In a scene-viewing experiment, we found that, for small objects, linguistics-based frequency, but not scene-based frequency, had effects on first fixation duration, gaze duration, and total time. Both linguistic and scene-based predictability affected total time. Similar to reading, fixation duration decreased with higher frequency and predictability. For large objects, we found the direction of effects to be the inverse of those found in reading studies. These results suggest that the recognition of small objects in scene viewing shares some characteristics with the recognition of words in reading.

Keywords: Scene viewing, word frequency, word predictability, LabelMe, Latent Semantic Analysis

Introduction

Eye Movements during Reading

It is well known that eye movements provide an indication of language processing because they are affected by lexical variables such as word frequency (the normative frequency of occurrence in a text corpus). Specifically, eye fixation times are longer on low-frequency words than on high-frequency words (Rayner, 1998). Rayner and Well (1996) also found that the predictability of target words has a strong influence on eye movements during reading. In their experiment, subjects fixated low-predictable target words longer than they did either high- or medium-predictable target words. Kliegl, Grabner, Rolfs, and Engbert (2004) also found the effect of word length, frequency, and predictability on inspection durations during reading.

Estimating Word Predictability during Reading

Typically, predictability is determined by the cloze task procedure (Taylor, 1953), in which subjects are asked to guess a word in a sentence from the prior sentence context. High-predictable words often reach probabilities of correct guessing between.70 and .90, whereas the probabilities for low-predictable words are below .10. Several computational alternatives, such as transitional probability (TP) by McDonald & Shillcock (2003), surprisal by Demberg & Keller (2008), co-occurrence probability (CCP) by Ong & Kliegl (2008) and Latent Semantic Analysis (LSA) by Wang, Pomplun,

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

Chen, Ko, & Rayner (2010) and Pynte, New, & Kennedy (2008) are used to predict eye movement behavior during reading.

LSA & Eye Movements in Reading

LSA is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text (Landauer & Dumais, 1997). To construct an LSA computation, a term-to-document matrix is first established from a corpus that embodies mutual constraints of semantic similarity of words. To solve these constraints, a linear algebra method, singular value decomposition (SVD), is applied to reduce the dimensions of the original matrix. The meaning of each word or passage is then represented as a vector in the resulting semantic space. LSA has been very successful at simulating a wide range of psycholinguistic phenomena, from judgments of semantic similarity to word categorization to discourse comprehension and judgments of essay quality (see Jones & Mewhort, 2007, for a review). It was tested on word predictability in the study by Wang et al. (2010), who reanalyzed the predictable/unpredictable target words from Rayner, Ashby, Pollatsek, & Reichle (2004). The results indicated that the predictable/unpredictable words determined by a cloze task can be distinguished by LSA. Wang et al. also suggested that LSA estimates higher-level lexical processing in reading because LSA influenced late processing measures, making it a complementary tool for deriving word predictability ratings.

Measures of Processing Time

To investigate eye movement behavior in reading, researchers typically use word-based measures such as across-subject averages of how often and for how long individual words are fixated. A number of word-based measures have become the standard (see Reichle, Rayner, and Pollatsek, 2003, for a review), including first fixation duration (FFD, the duration of the first fixation on a word independent of whether it is the only fixation on a word or the first of multiple fixations on it), gaze duration (GD, the sum of all fixation durations prior to moving to another word), and total time (TT, the duration sum of all fixations on a word including regressions).

Real-World Scene Viewing

The influence of semantic factors on fixation time during scene viewing has been studied (Loftus & Mackworth, 1978; Friedman, 1979; Antes & Penland, 1981; De Graef et al., 1990; Henderson et al., 1999; Hollingworth et al., 2003; Henderson & Ferreira, 2004; Võ & Henderson, 2009; see Nuthmann, Smith, Engbert, & Henderson, 2010, for a review). These studies have investigated the semantic consistency of a specific object with its scene context. The general finding is that total fixation times were longer for anomalous than for semantically consistent objects in scenes.

Estimating Object Predictability during Scene Viewing

It is important to notice that semantic consistency in the above studies is not equal to predictability as defined in reading studies. When computing predictability measures for words, subjects are given the sentence context and are asked to guess the next word. This assessment of how predictable a word is differs from determining how well the word fits into the sentence or context. Unlike words in a sentence, objects in a real-world scene do not line up from left to right and do not impose any particular sequence of processing. It is thus impossible to compute predictability in real-world scenes in a way that is entirely analogous to the one used in reading studies. Therefore, our study attempts to estimate predictability using LSA (see below for details) for all available objects in a scene, which is similar to estimating the semantic consistency of these objects with the scene. Although, due to the intrinsic difference between texts and scenes, this method cannot fully represent predictability as it is commonly computed in reading studies, it is a reasonable approximation that allows us to investigate the influence of semantic factors on fixation durations in scene viewing.

Motivation and Objective

It has not been tested whether frequency and predictability effects apply to scene viewing in ways analogous to reading. One of the reasons for this fact may be that analyzing eye movements using object-based measures in scene images requires object segmentation. However, the results of automated segmentation and labeling of images are still unsatisfactory for such a purpose. To overcome this problem, we used the freely available LabelMe image dataset (Russell, Torralba, Murphy & Freeman, 2008) containing a large number of scene images that were manually segmented into annotated objects. The locations of objects are provided as coordinates of polygon corners and are labeled by English words or phrases. These labels

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

allowed us to examine object fixations based on the frequency and predictability of objects as determined by linguistic analysis as well as visual scene analysis. The goal of this study was to contrast the results of both analyses with previous data from reading research for a first comparison of semantic factors influencing fixation duration in scene viewing and reading.

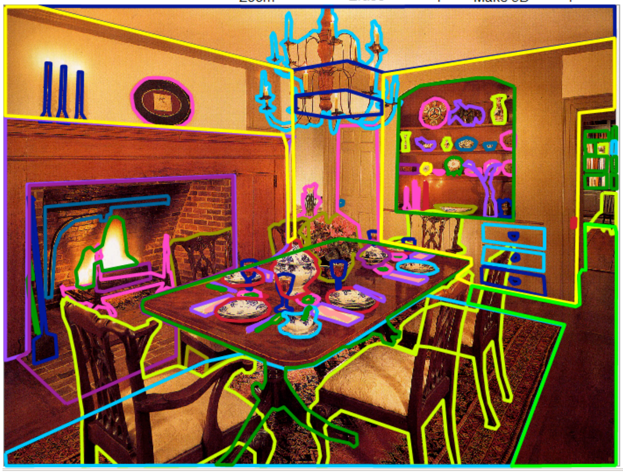

Figure 1. A dining room scene from the LabelMe database.

Methods

Participants

Twelve participants performed this experiment. All were students at the University of Massachusetts Boston, aged between 19 to 40 years old. Each participant received 10 dollars for participation in a half-hour session.

Apparatus

Eye movements were recorded using an SR Research EyeLink II system with a sampling frequency of 500 Hz. After calibration, the average error of visual angle in this system is 0.5°. Stimuli were presented on a 19-inch Dell P992 monitor with a refresh rate of 85 Hz and a screen resolution of 1024×768 pixels. Participants' responses were entered using a game-pad.

Materials

A total of 200 images (1024×768 pixels, 40°×30° of visual angle) of real-world scenes, including landscapes, home interiors, and city scenes, were selected from the LabelMe database (http://labelme.csail.mit.edu/, see Figure 1 for an example) as stimuli. Objects in each scene were annotated with corner coordinates of characteristic polygons defining the outline of the object shape and were labeled with English words. Each scene contained an average of 53.03 labeled objects (median = 40), covering 92.88% of the scene area.

Procedure

Following five practice trials, participants viewed the 200 scene images in random order. For each trial, they were instructed to inspect the scene and memorize it as thoroughly as possible. After a five-second presentation of each scene, an English word was shown for three seconds. Subjects had to manually report whether the object indicated by the word had been presented in the previously viewed scene. Target present and absent cases were evenly distributed among the 200 trials.

Data Analysis

Deriving Object Frequency and Predictability from Linguistic Analysis

Since the objects in the LabelMe scenes are labeled as English words or phrases, we are able to derive object frequency and predictability from a text corpus and LSA, similar to reading studies. Object frequency was computed from the British National Corpus (BNC).

For predictability, we used the "basic level" (Rosch and Mervis, 1975; Oliva & Torralba, 2001) of scene structure to describe each image in the materials, such as "landscape", "bedroom", "dining room", "office", "street", or "kitchen." Object predictability was estimated using the LSA@CU (http://lsa.colorado.edu/) tool with the semantic space "General Reading up to 1st year college (300 factors)." Object labels and scene gist descriptions in single English words were considered as "terms", while those in English phrases were considered as "documents" for LSA computation. For example, the object label "dish washer" and scene gist "kitchen" used "document to term" as comparison type, and the result was 0.42. Since LSA computes the cosine value between two vectors, the highest value in LSA computation is one. The cosine value is usually close to zero for random vector pairs in a high-dimensional space. Sometimes, the labels in the LabelMe dataset are not consistently applied to the same objects, for example, a "computer screen" in

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

one scene could be labeled "monitor" in another one. Since the cosine value of the vectors representing “computer screen” and “monitor” is high (0.6), we could still get similar results using LSA with different synonyms of labels. The resulting measure of predictability represents semantic consistency between an object and the gist of its embedding scene.

Deriving Object Frequency and Predictability from Visual Scene Analysis

Since the entire LabelMe database contains a large number of object labels, those labels can serve as a data source for computing frequency and predictability. The frequency of objects was accumulated across all available LabelMe scenes.

To compute predictability based on scene data, we established a semantic space using labels in LabelMe. There were 39,879 scene images and 303,033 annotated object labels (retrieved in February 2009). The objects with empty labels and the scene images without any object were excluded, which resulted in 39,724 scene images and 303,020 object labels. Since the labels were entered by a large number of contributors on the Internet, there existed different labels for identical objects (such as “car”, “SUV”, and “car occluded” for a car). We used a translating list (e.g. “car occluded” to “car”) provided by LabelMe to reduce the variability of object labels. This translation reduced the number of distinct object labels from 10,696 to 7,373.

We constructed a term-to-document matrix in which object labels served as terms and scene images served as documents. Subsequently, a term-to-document matrix containing 303,020 objects was established, and local weighting was performed (see Dumais, 1991). Local weighting is aimed at diminishing the influence of objects that are extremely frequent in an individual document. Often, researchers also apply global weighting to text corpora in order to reduce the importance of terms that occur in every document and therefore do not help to differentiate meaning, such as function words (“a” or “the”). However, we did not apply global weighting because the dataset of visual scenes was quite different from a typical text corpus. For example, object labels do not contain function words, which occur in every document in a text corpus. The computation of local weighting is described in Equation 1 (Dumais, 1991).

$$\textit{Local weighting} = \log(\text{term frequency} + 1) \qquad (1)$$

Subsequently, dimension reduction was performed on the term-to-document matrix, and a “semantic space” was established. In this semantic space, each vector had 500 dimensions, which is within the typical range for LSA studies (Landauer et al., 2007). To compute predictability, every object label and scene image was considered a vector in our LabelMe semantic space. We calculated the cosine value, representing semantic similarity, between the vectors of each object and its embedding scene image.

Identifying Fixated Objects

The proportion of the area in the selected scene images covered by annotated object regions was 92.88%. While this dense coverage is desirable for a comprehensive data analysis, it has the disadvantage that many objects were occluded by others. In fact, 34% of all fixations in the present study were located in the intersection of two or more object regions, making it difficult to identify the actually fixated, i.e., visible object that occluded the others. Even when we identified background objects by their labels (e.g., “WALL”, “FLOOR”, or “CEILING”), the percentage of fixations with multiple object regions was still 22%. Therefore, we estimated the depth-order of the intersecting objects based on the number of characteristic corners contributed to the intersection area by each object and the similarity of each object’s intersecting and non-intersecting parts in terms of their brightness (Histogram Intersection Similarity Method; Swain & Ballard, 1991).

Data Selection

There were a total of 19,767 objects fixated by all participants in our experimental data. Since the cognitive processes underlying the fixation of foreground and background objects might be different, we excluded all 1,512 cases in which background objects were fixated. We had to exclude another 677 cases because the labels of fixated objects were not included in the LSA@CU tool, resulting in a set of 17,578 remaining cases. Subsequently, we categorized these cases into high/low frequency, high/low predictability, and large/small object size groups by selecting the top and the bottom 6,000 cases for each variable. Only cases that were categorized by all three predictors (frequency, predictability, and size) were selected. This selection resulted in 5,816 cases in the linguistic analysis and 5,951 cases in the visual scene analysis.

(1)

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

Object Size

In reading studies, there is a clear relationship between the probability of fixating a word and its length: As the length of a word increases, so does its probability of being fixated (see Rayner, 1998, for a review). In scene viewing, large objects tend to receive more fixations than small ones. In our experiment, small objects contained an average of approximately 3,400 pixels, while large objects contained an average of approximately 90,000 pixels (see Table 2 below). As shown in Tables 3 and 4, large objects received longer gaze duration and total time than small objects. The large objects might often have been larger than the observers' perceptual span, and therefore multiple fixations were required to identify the object. In this study, we analyzed large and small objects separately.

Correlations among Predictors

The correlations among predictors acquired from linguistic (represented by suffix 'L') and visual scene (marked by suffix 'V') analyses are shown separately for small and large objects in Table 1. We found that FreqL (object frequency derived from BNC) and FreqV (object frequency accumulated from LabelMe) were highly correlated in both small and large objects. Both FreqL and FreqV were weakly correlated with PredL, which is consistent with reading studies. Moreover, both FreqL and FreqV were weakly correlated with PredV for small objects, but were somewhat correlated in large objects. This correlation might have been caused by not applying global weighting in the LabelMe semantic space. Therefore, although the influence of a highly frequent object in one scene was reduced by local weighting, highly frequent objects distributed across many scenes still received high predictability values. In addition, we found that Size was only weakly correlated with FreqL, FreqV, PredL, and PredV for either small objects or large objects, which means that object size did not significantly influence other predictors.

Eye Movement Analysis

We separated the raw data into four groups: linguistic analysis for small objects, visual scene analysis for small objects, linguistic analysis for large objects, and visual scene analysis for large objects. The data in these four groups were submitted separately to an analysis of variance (ANOVA) with frequency (low vs. high) and predictability (low vs. high) as within-subject factors. The average area covered by objects (number of pixels), the natural logarithm of frequency, and the cosine values indicating predictability of objects in each group are shown in Table 2.

Table 1
Correlation among predictors for small vs. large objects

	FreqL	FreqV	PredL	PredV	Size
FreqL	—	.494	-.032	.118	-.095
FreqV	.551	—	.068	.141	.155
PredL	-.013	.049	—	.108	.073
PredV	.340	.329	.149	—	.050
Size	.174	.036	.108	-.033	—

Note: The data for small and large objects are shown above and below the diagonal, respectively. FreqL is the natural logarithm of frequency from the British National Corpus (BNC); FreqV is the natural logarithm of frequency accumulated from LabelMe; PredL is the cosine value between the target object and its scene gist computed by LSA@CU; PredV is the cosine value between the target object and its scene computed for the LabelMe semantic space; Size is the number of pixels enclosed in the polygon of an object provided by the LabelMe dataset.

Table 2
The average area, frequency, and predictability of objects

			Frequency		Predictability	
	Size	Pixels	Low	High	Low	High
Ling	Small	3,350	6.24	9.90	0.01	0.39
Ling	Large	99,504	6.56	9.88	0.02	0.42
Visual	Small	3,495	4.21	9.67	0.30	0.66
Visual	Large	89,134	4.45	9.75	0.28	0.68

Note: Ling stands for linguistic analysis, and Visual stands for visual scene analysis; Pixels is the number of pixels enclosed in the polygon of an object provided by the LabelMe dataset.

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

Results and Discussion

Linguistic Analysis – Small Objects

First Fixation Duration (FFD). In this study, FFD is defined as the first fixation on an object, regardless of whether it is the only fixation on an object or the first of multiple fixations on it. The main effect of frequency was found to be significant, F(1, 11) = 5.60, p<.05, indicating that high-frequent objects received less fixation time than low-frequent ones. There was neither a main effect of predictability, F(1, 11) < 1, nor an interaction of the factors, F(1, 11) = < 1. The results of the FFD analysis in terms of mean values and their standard deviation are shown in Table 3.

Table 3
Summary of mean and standard deviation (in ms) based on linguistic analysis

Measurement	Size	Freq	Pred	Mean	Std
FFD	Small	Low	Low	265	29
			High	271	44
		High	Low	260	41
			High	252	29
	Large	Low	Low	260	35
			High	269	37
		High	Low	260	46
			High	268	34
GD	Small	Low	Low	302	36
			High	301	52
		High	Low	288	40
			High	285	45
	Large	Low	Low	390	62
			High	431	71
		High	Low	444	111
			High	505	109
TT	Small	Low	Low	340	45
			High	334	62
		High	Low	333	49
			High	314	56
	Large	Low	Low	496	79
			High	576	102
		High	Low	582	127
			High	694	120

In reading studies, FFD usually reflects early visual and lexical processing (including identification of orthographic form and a familiarity check), and is affected by both frequency and predictability (see Rayner, Ashby, Pollastsek, & Reichle, 2004, for a review). Consistent with reading studies, during scene viewing, we observed a frequency effect. However, in contrast to reading, we failed to obain a predictability effect. Two factors may be responsible for this pattern of results: First, we suggest that high-frequent objects might be processed faster than low-frequent objects, which is similar to the fact that high-frequent words are processed faster than low-frequent words. Second, since predictability was estimated as object-scene consistency, predictability effects are likely to only affect late stage processing, which might be reflected in total time (TT) but not FFD.

Gaze Duration (GD). In scene viewing, first-pass GD is hard to compute because there is no default direction of visual scanning such as the first left-to-right sweep over each sentence in English reading. In addition, due to the different sizes and shapes of objects, it is also difficult to define a "spotlight" and determine which objects in the scan path were actually processed. Therefore, we computed GD using the sum of all fixation durations prior to moving to another object. We found a significant effect of frequency, F(1, 11) = 10.00, p<.01, but not for predictability, F(1, 11) < 1. The interaction of frequency and predictability was not significant either, F(1, 11) < 1. The mean values and their standard deviation are shown in Table 3.

In reading studies, GD reflects lexical processing, such as the identification of a word's phonological and/or semantic forms, and lexical access. GD is also often influenced by both frequency and predictability during reading (see Rayner et al., 2004, for a review). The current frequency effect, and the fact that it is more pronounced for GD than for FFD, is consistent with findings from reading studies. We presume that the frequency effects were due to the same reasons as for FFD.

Total Time (TT). In this study, TT is computed as the duration sum of all fixations on an object including regressions. The main effects of frequency and predictability on TT were significant, F(1, 11) = 6.92, p<.05, and F(1, 11) = 5.78, p<.05, respectively, while there was no

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

significant interaction of frequency and predictability, F(1, 11) < 1. The mean and standard deviation of TT across factors are shown in Table 3.

In reading research, TT includes gaze duration and refixations and is thought to reflect information integration. In the present study, both frequency and predictability effects on TT were found. We suggest the explanation that participants re-fixate low-frequent or low-predictable objects more often than high-frequent or high-predictable objects. Both effects were consistent with results from reading studies.

Linguistic Analysis – Large Objects

First Fixation Duration. The main effect of predictability on FFD was significant, F(1, 11) = 4.96, p<.05; surprisingly, high-predictable objects received more fixation time than low-predictable ones. This predictability effect is the inverse of that found in reading studies. There was neither a main effect of frequency, F(1, 11) < 1, nor an interaction of the factors. The results of the FFD analysis in terms of mean values and their standard deviation are shown in Table 3.

Gaze Duration. We found significant main effects on GD for both frequency, F(1, 11) = 13.35, p<.01, and predictability, F(1, 11) = 10.75, p < .01. High-frequent and high-predictable objects were fixated longer than low-frequent and low-predictable ones, respectively, which is the inverse of results from reading studies. The interaction was not significant. The mean values and their standard deviation are shown in Table 3.

Total Time. The main effects of frequency and predictability on TT were significant, F(1, 11) = 24.44, p<.001, and F(1, 11) = 23.14, p<.001, respectively. The direction of the effects was identical to GD and, again, the inverse of that found in reading studies. As in the GD analysis, there was no significant interaction of frequency and predictability for FFD, F(1, 11) < 1. The mean and standard deviation of TT across factors are shown in Table 3.

Visual Scene Analysis – Small Objects

First Fixation Duration. We failed to obtain significant effects of frequency, predictability, or their interaction on FFD, all Fs(1, 11) < 1. These results suggest that frequency and predictability derived from visual scene analysis might not be good predictors of FFD.

Gaze Duration. In scene viewing, again, we do not obtain frequency, predictability, or interaction effects on GD, F(1, 11) < 1, F(1, 11) = 2.67, and F(1, 11) < 1,all ps > .1. These results were similar to FFD.

Total Time. The main effect of predictability on TT was significant, F(1, 11) = 9.88, p < .01; low-predictable objects received more total time than high-predictable objects. There was no frequency effect, F(1, 11) = 2.36, p > .1 and a marginal interaction, F(1, 11) = 4.59, p = .055. These results suggest that, similar to linguistic analysis, predictability tends to affect TT, which reflects information integration. The mean and standard deviation of TT across factors are shown in Table 4.

Visual Scene Analysis – Large Objects

First Fixation Duration. Using the visual scene measure, we failed to obtain significant effects of frequency, predictability, or their interaction on FFD for large objects, F(1, 11) = 2.55, F(1, 11) = 1.80, and F(1, 11) < 1, respectively, all ps > .1.

Gaze Duration. Similarly to FFD, there were no effects of frequency, predictability, or their interaction on GD, F(1, 11) < 1, F(1, 11) = 2.06, and F(1, 11) < 1, all ps > .5.

Total Time. We found a marginal effect of predictability on TT, F(1, 11) = 4.52, p = .057, indicating a trend for high-predictable objects toward receiving greater TT than low-predictable objects. The effect was the inverse of what has been found in reading studies. There was no frequency effect, F(1, 11) < 1, and no interaction, F(1, 11)

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

< 1. The mean and standard deviation of TT across factors are shown in Table 4.

Table 4
Summary of mean and standard deviation (in ms) based on visual scene analysis

Measurement	Size	Freq	Pred	Mean	Std
FFD	Small	Low	Low	267	40
			High	277	34
		High	Low	275	93
			High	272	41
	Large	Low	Low	269	32
			High	280	40
		High	Low	255	42
			High	268	28
GD	Small	Low	Low	321	64
			High	310	41
		High	Low	332	95
			High	297	40
	Large	Low	Low	442	68
			High	464	131
		High	Low	417	86
			High	460	89
TT	Small	Low	Low	362	77
			High	353	52
		High	Low	421	99
			High	336	42
	Large	Low	Low	587	83
			High	636	190
		High	Low	560	96
			High	617	95

General Discussion

The results for small objects indicate that FreqL has effects on FFD, GD, and TT that are similar to those found in reading studies. Although the correlation of FreqL and FreqV is high (see Table 1), the results suggest that FreqL is a better predictor of FFD, GD, and TT compared to FreqV.

It is interesting that although PredL and PredV were only weakly correlated, both PredL (object predictability derived from linguistic analysis) and PredV (object predictability computed from visual scene analysis) influenced TT. We suggest that both PredL and PredV capture object-scene consistency, which in both cases influenced TT. The frequency effects observed in scene viewing on FFD might reflect early visual processing, those on GD might correspond to higher-level cognitive activities (such as semantic activation), and those on TT might be due to information integration as observed in reading studies. Predictability effects on TT in scene viewing might reflect late-stage semantic verification of object-scene consistency.

Object size had a substantial influence on GD and TT - larger objects were fixated longer. More importantly, small and large objects induced very different frequency and predictability effects: For large objects, frequency effects were found on GD and TT in the linguistic analysis but not in the visual scene analysis. Predictability effects were found on FFD, GD, and TT in the linguistic analysis and on TT in the visual scene analysis. Interestingly, the direction of all effects for large objects was the inverse of that found in reading studies.

We suggest that the processing of large objects might be particularly demanding and thus induce gaze behavior that is substantially different from processing both small visual objects and written words. Conceivably, the size and complexity of large objects may often not allow its inspection within a single fixation. In the current memorization task in which the participants need to quickly develop a semantic understanding of the scene, inspecting objects that are frequent and predictable (i.e., consistent with the scene gist) may be most efficient. Therefore, the inspection of large objects that are less useful in this regard may not be completed but interrupted after the initial fixation. If such effects exist, it is plausible that the object-based measures for large objects in scene viewing play a different role than those for both small objects and written words.

Based on the current results, we propose that frequency and predictability, from both visual scene and linguistic analysis, and size should be taken into account to develop a computational model of fixation durations in scene viewing. As discussed above, it is important to notice that the current same-direction effects for small objects and inverse effects for large objects were observed in a brief-presentation memorization task. We suggest that such effects may vary for different tasks (for example, long-presentation memorization tasks, visual search, object counting, or scene gist recognition), which we will investigate in future studies. Regarding the design of such

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

experiments, the current data indicate that LabelMe and LSA are useful, complementary tools for studying eye movements during scene viewing.

Acknowledgement

Thanks to Keith Rayner and Reinhold Kliegl for helpful comments on an earlier version of the article. Preparation of the article was partially supported by NIH grant R15EY017988 to M.P. Correspondence should be addressed to Hsueh-Cheng Wang (hchengwang@gmail.com).

References

Demberg, V. & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*. 109, 193-210.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. *In J. M. Henderson and F. Ferreira (Eds.), The interface of language, vision, and action: Eye movements and the visual world* (pp. 1-58). New York: Psychology Press.

Jones, M. N. & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114, 1-37.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262-284.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review,* 104, 211–240.

Landauer, T. K., McNamara, D. S., Dennis S., & Kintsch W. (2007). *Handbook of Latent Semantic Analysis*, Lawrence Erlbaum Associates.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science,* 14, 648–652.

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, 117(2), 382-405.

Oliva, A. & Torralba, A. (2001). Modeling the Shape of the Scene: A holistic Representation of the Spatial Envelop. *International Journal of Computer Vision*, 42 (3), 145-175.

Ong, J. K. Y. & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research*, 2(1):3, 1-7

Pynte, J., New, B. & Kennedy, A. (2008). A multiple regression analysis of syntactic and semantic influences in reading normal text. *Journal of Eye Movement Research*, 2(1):4, 1-11.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin,* 124, 372-422.

Rayner, K. (2009). The Thirty Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457-1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance,* 30, 720–732.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review,* 3, 504–509.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences,* 26, 445–476.

Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605.

Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. (2008), LabelMe: a database and web-based tool for image annotation, *International journal of computer vision,* 77, 1-3, 157-173.

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, 9(3):24, 1-15.

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692

Wang, H. C., Pomplun, M., Ko, H. W., Chen M. L., & Rayner, K. (2010). Estimating the effect of word predictability on eye movements in Chinese reading using latent semantic analysis and transitional probability, *Quarterly Journal of Experimental Psychology*, 63, 1374-1386.

DOI 10.16910/jemr.3.3.3

ISSN 1995-8692