Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Semantic Processing Modulates the Attentional Accessibility of Verbal and Nonverbal Search Targets

Permalink

https://escholarship.org/uc/item/59p6h6kp

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Mauti, Marika
D'Innocenzo, Giorgia
Buonocore, Antimo
et al.

Publication Date

2024

Semantic Processing Modulates the Attentional Accessibility of Verbal and Nonverbal Search Targets

[1]Marika Mauti[1*], Giorgia D'Innocenzo[2], Antimo Buonocore[3], Moreno I. Coco[1,4*]
[1]Department of Psychology, Sapienza University of Rome, Rome, Italy
[2]CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal
[3]Department of Educational, Psychological and Communication Sciences, Suor Orsola Benincasa University, Naples, Italy
[4]I.R.C.S.S. Santa Lucia, Rome, Italy

Abstract

The seminal dual coding theory by Paivio (1971) posited that non-verbal and verbal stimuli differ in their representational format, whereby the former activates a dual code while the latter only one. These differences in code have implications for tasks such as visual search. The current eye-tracking visual search study aims to re-evaluate this theoretical framework while examining the role played by semantic processing that has never been looked at before. We followed the original design by Paivio and Begg (1974), with participants searching for a target, cued either by a word or a picture, in an array of either words or pictures. The target could be either semantically related or unrelated to the other distractors. Corroborating original results, response times for correct trials were faster in pictorial arrays and substantially slower when a cued picture had to be found in a word array. Semantically unrelated targets were looked at faster for longer, leading to shorter search responses than semantically related targets. Critically, these effects driven by semantic relatedness were amplified when codes had to be converted (e.g., picture-to-word). Our findings refine our understanding of the role semantic processing plays in the representational format of words and pictures and the implications it carries for visual search.

Keywords: Object semantics; Extrafoveal processing; Dual Coding Theory; Visual search; Eye movements

Introduction

Visual search is essential to most daily tasks, from discriminating the car keys from a cluttered desk to browsing verbal information on crowded websites. Critically, it is still debated what visual information gets acquired across the visual field and the difference it may make if such information is verbal or non-verbal. Seminal research on the search for non-verbal (i.e., pictorial) targets suggested a two-stage process whereby low-level information (e.g., colour, shape and orientation) is acquired immediately and in parallel, across the visual field (i.e., extra-foveal) while in a subsequent stage, high-level information (e.g., semantics) needs foveal vision to be acquired (see Treisman and Gelade, 1980 for classic empirical work and Wolfe, Cave, and Franzel, 1989 for an early model, i.e., Guided Search). More recently, but with a similar perspective, Zelinsky (2008) also suggested a two-stage process, the first guided by the perceptual memory template (low-level) of the target to be found, followed by a second step involving the semantic recognition of the target to finalise its discrimination from other visual distractors. Differently from what these visual search models predict, it became clearer more recently that semantic information can also be quickly extracted in extra-foveal vision and used to guide the early allocation of overt attention (Belke et al., 2008; Cimminella et al., 2020; Nuthmann et al., 2019, and see Coco et al., 2020; LaPointe & Milliken, 2016 for naturalistic scenes). When considering verbal information (e.g., words), the extent to which semantic information is acquired from the parafovea remains more uncertain (Andrews & Veldre, 2019). Reading studies on German (Hohenstein & Kliegl, 2014), Italian (Rusich et al., 2020) and Chinese speakers (Yan et al., 2009) suggest that semantics can be processed in parafovea even though such results do not seem to extend to English (e.g., Rayner, Balota, & Pollatsek, 1986). In search tasks involving words as targets, overt attention is attracted by distractor words sharing orthographic, phonological, and semantic features with the target (Dampuré et al., 2014; Leger et al., 2012). Still, acquiring visual and semantic information in parafovea decreases with increasing cognitive load (Dampurè et al., 2019) and improves with a low foveal load (Antùnez et al., 2022). As written words link visual orthographic codes with a conceptual representation of their meaning, they may not be searched using the same attentional strategies deployed for visual objects, which are instead more explicit in the mapping between perceptual and conceptual information. The idea that words and pictures have different representational formats can be traced back to the dual coding theory (DCT; Paivio, 1971), which states that words are encoded through a single verbal code, while pictures have a dual code, visual and verbal, which make them easier to access and more memorable. Paivio and Begg (1974) were the first to test the implication of this coding difference for visual search. In their experiment, participants had to search and point at a target, cued either as a word or a picture, in an array of distractors that were again words or pictures. Across the board, search times were faster for picture arrays than word arrays, but picture cues facilitated search only in a picture array. These findings indicate that the dual code of pictures may help their accessibility, but a conversion cost arises when pictures must be searched into a word array. This

---

[1] *Correspondence about this work can be addressed to Marika Mauti (marika.mauti@uniroma1.it) and Moreno Coco (moreno.coco@uniroma1.it)

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

seminal study only collected manual behavioural responses, so remaining agnostic as to how overt attention would be allocated during the search. To the best of our knowledge, the study by Hurley et al. (2021) was the only one aiming to replicate Paivio and Begg (1974) while looking at overt attention (i.e., eye movements) and focusing on the hypothesis that the search for pictorial information is more efficient than for words because it relies on parallel as opposed to serial processing. They observed more fixations to distractors and smaller saccade amplitudes when the search was conducted within a word array compared to a picture array. The interpretation of these findings was that locating a word requires serial processing compared to pictures that can be detected through parallel processing. Even if this explanation of the picture superiority is compelling, Hurley et al. (2021) did not examine how semantic processing may operate when verbal and pictorial codes must be converted (e.g., a search for a pictorial target in a word array).

The current eye-tracking study precisely aims to fill this gap by investigating the role played by semantic processing in the extrafoveal capture of overt attention in a visual search task manipulating the verbal and non-verbal nature of the target and its embedding context. In line with Paivio and Begg (1974), we expect faster response times for pictorial arrays than verbal ones, independently of cue modality. We also expect to replicate effects driven by the semantic relatedness between the target and the other distractors in the array. We predict that targets semantically unrelated to the distractors will be prioritised in early attention and associated with faster response times, shorter latency of first fixation and longer dwell times (e.g., Nuthmann et al., 2019; Cimminella et al., 2020). When reflecting upon the impact of code type and on those cases when the conversion is required, i.e., when a cue is in one modality (e.g., picture) but the array in the other (e.g., word), we expect to replicate the cost observed by Paivio and Begg (1974) when a picture needs to be found in a word array. We posit that converting from a cue word to a picture array is faster because it only activates the orthographic concepts of the word while searching for a picture is facilitated by its dual code (i.e., perceptual and conceptual). Conversely, when a picture needs to be searched in a word array, its perceptual aspect is not necessary, and it could hamper the search, resulting in a slowdown of RTs. Furthermore, searching for a picture in an array should be faster because it is primarily guided by the perceptual appearance of the target (e.g., Zelinksy, 2008). There should be no difference due to the semantic relatedness of the target with the other distractors, as the search should be primarily guided by its perceptual appearance. However, suppose semantic information is accessed regardless of the guidance provided by the perceptual template matching. In that case, we should observe a prioritisation (e.g., faster RTs) for targets semantically unrelated to other distractors. We expect this reasoning to apply to the scenario where the target is not present in the array but is replaced with a similar critical object, so its perceptual template cannot be used to guide attention. For example, cueing the picture of a *ladybug* when the critical object is a *bee* would lead to a faster exclusion of other distractors if semantically unrelated (e.g., *vegetables*).

Methods

Participants

Thirty undergraduate students (22 females; age = 22.3 ± 5.6) Portuguese native speakers enrolled at Universidade de Lisboa with normal or corrected-to-normal vision took part in the study and received one credit as compensation for their time. The local Ethics Committee approved the study before commencing data collection.

Design

A 2x2x2x2 factorial within-participant design was implemented by crossing Cue Modality (picture, word), Array Modality (picture, word), Semantic Relatedness (related, unrelated), and Target (present, absent).

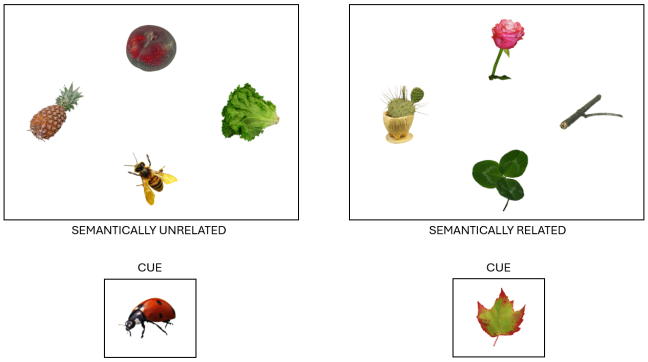
Stimuli

Stimuli comprised 488 images of objects from the Bank of Standardized Stimuli (BOSS) database (Brodeur et al., 2010, 2014). The Portuguese names for the visual objects were obtained using a brief survey implemented on Qualtrics (Provo, UT, USA; https://www.qualtrics.com). Ninety-three native Portuguese speakers (72 female, age = 31.5 ± 11.32) were presented with an object per trial and asked to type in its corresponding name, and in case of not knowing, input “do not know." Objects that contributed to creating the experimental arrays had a minimum of 70% naming agreement across participants and were less than 20% classified as "do not know". When participants used multiple names to refer to the visual object, only the modal name was retained (e.g., "stick de cola," "tubo de cola," and "batom de cola" were all scored as "cola"). Plural/singular and masculine/feminine variations of the same response were standardised by selecting the most frequent response (e.g., if "escadas" was the most frequent response, "escada" was transformed into its plural form). After this first selection, objects had to be assembled into arrays (four visual objects or four written words each) and manipulated based on their semantic relatedness. Semantically unrelated arrays were arranged to have three distractors from the same semantic category (e.g., vegetables) and a target from a different category (e.g., animals; see Figure 1 for a visualisation). In semantic-related arrays, all objects, including the target, belonged to the same category (e.g., all plants). Although the perceptual similarity of the stimuli has not been manipulated, given the absence of systematicity among low-level features (e.g., shape and colour) between the cue and the objects within the array, we expect that the study results would be confirmed even by removing perceptual properties of the objects, for instance using greyscale images.

We made sure that the target name was balanced in the related and unrelated condition for the number of characters (Unrelated = 5.91 ± 1.41; Related = 5.92 ± 1.39; t = -0.29; p = 0.7) and lexical frequency (Unrelated = 2.66 ± 0.69;

Related = 2.68 ± 0.68; t = -0.6; p = 0.5). We had 112 unique arrays, of which 56 pictures and 56 words. Each array was associated with a cue, which was again either a word or a picture, which resulted in four possible experimental combinations: (a) picture cue to picture array, (b) word cue to picture array, (c) picture cue to word array, and (d) word cue to word array. In our experimental design, trials might be target-present or absent. In target-present trials, the critical object was always the cued target. In contrast, in target-absent trials, a critical object was cued as a different item (word or picture) which belonged to the same semantic category as the target. So, for example, in the semantically related trials, participants could be cued with a *leaf.* However, the *rose* is the critical object in the array (among other *plants*). In contrast, for semantically unrelated arrays, the cue is a *ladybug,* and the critical object is a *bee* among vegetables. Despite being target-absent trials, one of the four distractors was selected a priori in each array as the "target", which, through random rotation, occupies a different position (top, down, left, right) in each of the four repetitions of the array. This enabled the evaluation of the semantic effect on RTs and eye movement measures. In the target-absent semantically related condition, the a priori selection of a critical object as target controlled and balanced potential spatial bias that may influence our variables of interest. Each word and each picture in the array had a resolution of 250 x 250 pixels and were placed such that had a distance of ~7.5 cm from the centre of the screen in the four cardinal directions, which corresponds to 7 degrees of visual angle at a viewing distance of 60 cm and guarantees that their position was in the extrafoveal section of the visual field (see Figure 1 for a visualisation).

Figure 1. An example of semantic relatedness manipulation for the condition where the cue is a picture, the array is also made of pictures, and the search target is absent. Note that word cues or word arrays were constructed by simply substituting the pictorial objects with the modal names obtained through norming.



Apparatus and Recording

Stimuli were presented with a 21-in. monitor (LCD DELL 1920 × 1080 px) at a refresh rate of 60 Hz. Eye movements were recorded binocularly using a LiveTrack Lightning tracker (Cambridge Research Systems) at a sampling rate of 500 Hz. Still, analyses were conducted on data from the dominant eye as determined with a parallax test. A chin and forehead rest were used to stabilise the participant's head. A nine-point calibration was run at the beginning of each session (visual angle error; x-axis = 0.66° ± 0.83; y-axis = 0.73° ± 1.09 SD). Electroencephalography (EEG) was concurrently recorded from 64 active electrodes at a sampling rate of 512 Hz using BioSemi ActiveTwo amplifiers. Six electrodes are located near the left and right canthus and above and below both eyes to record the Electrooculography (EOG), and two are placed on the left and right mastoid. Eye tracking and EEG data were synchronised using shared triggers sent via the parallel port of the stimulus presentation PC to the two recording computers. We report that EEG responses were also collected for completeness and scientific rigour. However, we will only analyse and discuss results related to the manual behavioural and eye movement responses in the current study due to space and focus. The experiment was implemented on MATLAB (Version R2021a) using the Psychtoolbox extension (Version 3.0.19) (Kleiner et al, 2007).

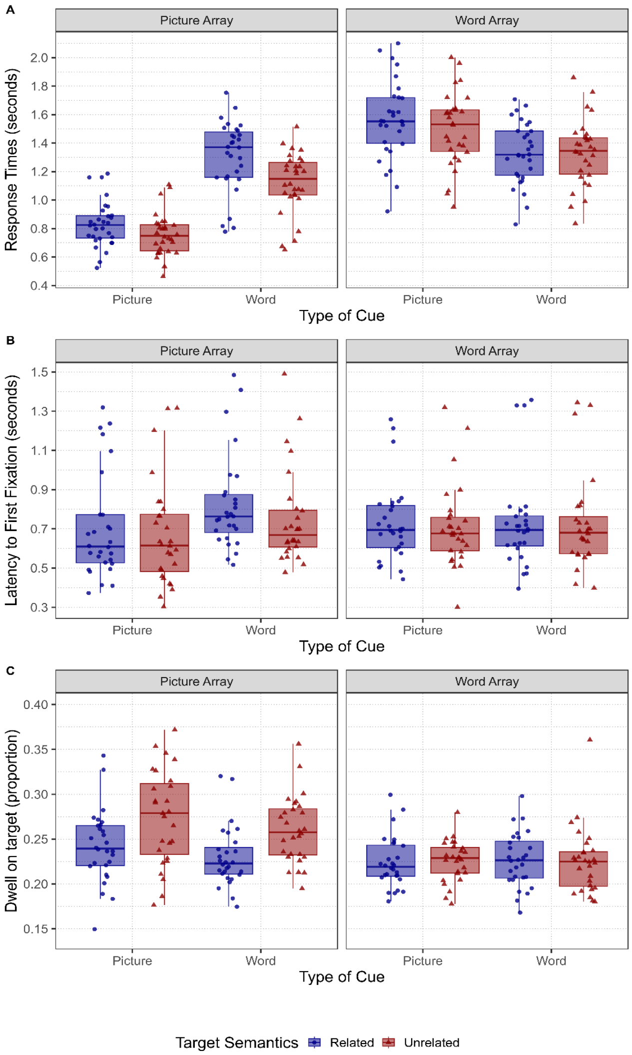
Procedure

Participants were asked to read and sign a consent form and then received written instructions about the task at the start of the experimental session. Then, the eye tracker and EEG were set and calibrated. Participants were seated at 60 cm from the computer screen. Each search trial began with a cue (either a word or a picture) presented for 800 ms, followed by a fixation cross with a duration of 250 ms, and then the array was presented. Participants had to press the keyboard to indicate whether the target was (or not) present (“a” and “l”, respectively). They had a timeout of 6 seconds, after which a null response was logged. Each participant completed four blocks (picture cue to picture array, word cue to picture array, picture cue to word array, word cue to word array) comprising 56 trials each. All four blocks were repeated four times, and their order of presentation was randomised before each repetition. Each participant performed 896 trials (half of these were target-absent trials). The position of the critical object systematically rotated in each repetition, covering the four cardinal positions (i.e., top, left, bottom, right). The repetitions were implemented to significantly increase the data points for each experimental condition, which will be necessary for future analyses of fixation-related potentials. Thus, each participant completed 448 target-absent trials, which were equally balanced across conditions (i.e., 56 * 4 repetitions per condition). After completing 50% of the experimental trials, participants were given a 5-minute break, and the eye tracker recalibrated just after it. The experimental sessions lasted approximately 2 hours.

Data analysis

Our analyses will focus on trials where the target is absent. This choice is driven by the need to (i) streamline the structure of our inferential models and (ii) investigate the influence of semantic relatedness, regardless of the effects of working memory, in instances of target-present trials (i.e., search guidance driven by the target template[2]; and see Belke et al. 2008 or Cimminella, et al., 2020, showing stronger effects of semantic relatedness in target-absent trials). *Response times* of correct responses were considered for the inferential analysis. Saccade and fixation events were detected from the raw x and y coordinates samples using the binocular version of the velocity-based microsaccade detection algorithm by Engbert & Kliegl (2003) as implemented in the EYE-EEG toolbox in MATLAB (velocity threshold: 6 median-based SDs, minimum saccade duration: 8 ms). For saccades separated by less than 25 sample fixations, only the first saccade is kept). Two participants had to be excluded from the eye-movement analysis as the overall quality of the eye-tracking data was poor. Fixation coordinates were then mapped using rectangular Areas of Interest (AOIs) surrounding the critical object and used to extract two dependent measures: (a) *Latency to First Fixation*, which represents the time between the onset of the search array and the first fixation on the critical object and points at parafoveal processing (i.e., the shorter the latency, the more likely the object was prioritised in early attention) and (b) *Dwell Time*, which is the proportion of time fixating onto the critical object and points at the foveal effort to acquire information about it (i.e., the greater the proportion, the more the effort). Since we analysed data from the target-absent condition here, both eye movement measures refer to the a priori designated critical object (see Figure 1 for a graphical representation of the experimental manipulation). Statistical inference was obtained with Generalized Linear-effect Models (G/LMER) implemented in the `lme4` package in R (Bates et al., 2015). We predict the three dependent measures above discussed separately for picture and word arrays as predicted by (i) *Semantic Relatedness* (related and unrelated, with related as the reference level), (ii) *Repetition* (as a continuous variable from 1 to 4) and (iii) *Cue Modality* (word and picture, with picture as the reference level). The random variables considered are Participant (30 for manual responses and 28 for eye movement responses) and Item (56). Models were first built with complete fixed and random effect structure, i.e., all main effects and interactions, introducing the random variables both as intercept and slopes (Barr et al., 2013), then reduced backwards using the step function from the `lmerTest` package (Kuznetsova et al., 2017) to obtain the final model with the most parsimonious number of parameters that best fitted the data (Matuschek et al., 2017). Pairwise comparisons were obtained using the emmeans package and directly reported in the text to establish the direction of effects in categorical interaction terms. In the Tables, we report the coefficients of the predictors retained in the final model, the confidence intervals to judge the uncertainty in our estimates (see Luke, 2017) their t-values and mark the p-values with asterisks, based on asymptotic Wald tests computed using the `lmerTest` package.

Figure 2. Box plots of the response times (A; RTs), latency to first fixation (B) and proportion of dwell time on the critical object (C) divided by semantic relatedness within each panel (blue, related; red, unrelated) for pictures and word arrays arranged as left and right panels respectively. The hinges of the boxplots represent the 25th and 75th percentiles of the measure (lower and upper quartiles), while the horizontal line represents the median of the distribution. Each dot indicates the by-participant average for that measure.



---

[2] Analysis of target present trials corroborates the results from target-absent trials, even if semantic relatedness effects have a reduced magnitude.

Results

For analyses focusing on the picture array, we found a significant main effect of semantic relatedness, with shorter RTs for unrelated than related objects, especially when participants were cued by a word [pairwise-comparison; *z-ratio* (9.66), p < .0001] (refer to Figure 2A, left panel and Table 1 for the model coefficients). Repetition was also a significant main effect whereby RT decreased for increasing repetitions, again, especially for word cues. The latency to first fixation was also significantly faster for a semantically unrelated critical object but only when the cue was a word [pairwise-comparison; *z-ratio* (3.56), p < .05] (Figure 2B, left panel).

Table 1. Generalized Linear Mixed-effects Model for RTs, Latency to First Fixation and Dwell Time of the picture array condition as predicted by Semantic Relatedness (with semantically related critical object as the reference level), Repetition (continuous variable from 1 to 4) and Cue Modality (word, picture, with picture as the reference level). The random variables introduced as intercept and slopes were Item (56) and Participant.

<table>
<tr><th colspan="4">Reaction Times (RTs)</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>0.89</td><td>0.81;0.96</td><td>23.76***</td></tr>
<tr><td>Semantic Relatedness</td><td>-0.08</td><td>-0.12;-0.04</td><td>-3.78***</td></tr>
<tr><td>Repetition</td><td>-0.02</td><td>-0.04;-0.003</td><td>-2.34*</td></tr>
<tr><td>Cue Modality</td><td>0.69</td><td>0.63;0.76</td><td>20.35***</td></tr>
<tr><td>Semantic Rel.: Cue Mod.</td><td>-0.12</td><td>-0.15;-0.08</td><td>-6.57***</td></tr>
<tr><td>Repetition: Cue Mod.</td><td>-0.08</td><td>-0.10;-0.07</td><td>-10.64***</td></tr>
<tr><th colspan="4">Latency to First Fixation</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>0.69</td><td>0.56;0.81</td><td>10.33***</td></tr>
<tr><td>Semantic Relatedness</td><td>0.005</td><td>-0.13;0.15</td><td>0.07</td></tr>
<tr><td>Repetition</td><td>0.001</td><td>-0.04;0.04</td><td>0.07</td></tr>
<tr><td>Cue Modality</td><td>0.27</td><td>0.15;0.39</td><td>4.48***</td></tr>
<tr><td>Semantic Rel.: Repetition</td><td>-0.01</td><td>-0.07;0.037</td><td>-0.55</td></tr>
<tr><td>Semantic Rel.: Cue Mod.</td><td>-0.24</td><td>-0.41;-0.07</td><td>-2.76**</td></tr>
<tr><td>Repetition: Cue Mod.</td><td>-0.05</td><td>-0.09;-0.008</td><td>-2.30*</td></tr>
<tr><td>Semantic Rel.: Rep.: Cue Mod.</td><td>0.07</td><td>0.011;0.014</td><td>2.28*</td></tr>
<tr><th colspan="4">Dwell Time</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>0.23</td><td>021;0.25</td><td>25.89***</td></tr>
<tr><td>Semantic Relatedness</td><td>0.03</td><td>0.02;0.04</td><td>6.87***</td></tr>
<tr><td>Repetition</td><td>0.005</td><td>0.001;0.009</td><td>2.35*</td></tr>
<tr><td>Cue Modality</td><td>-0.02</td><td>-0.03;-0.00</td><td>-3.35**</td></tr>
</table>

(*) p < .10, *p < 0.05, **p < .01, ***p < .001

In general, the critical object was looked at later when the cue was a word as compared to a picture. Also, the dwell time was significantly influenced by semantic relatedness with participants fixating for longer an unrelated compared to a related critical object (Figure 2C, left panel). With increasing repetitions, the dwell time of the critical object also increased. Finally, the critical object was looked at less when cued by a word. For the analyses focusing on the word array, we observed a significant main effect of semantic relatedness on RTs with shorter RTs for unrelated compared to related critical objects, which is qualified by a significant interaction with the modality of the cue (refer to Figure 2A, right panel and Table 2 for the model coefficients).

Table 2. Generalized Linear Mixed-effects Model for RTs, Latency to First Fixation and Dwell Time of the word array condition as predicted by Semantic Relatedness (with semantically related critical object as the reference level), Repetition (continuous variable from 1 to 4) and Cue Modality (word, picture, with picture as the reference level). The random variables introduced as intercept and slopes were Item (56) and Participant.

<table>
<tr><th colspan="4">Reaction Times (RTs)</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>1.77</td><td>1.64;1.90</td><td>27.70***</td></tr>
<tr><td>Semantic Relatedness</td><td>-1.10</td><td>-0.15;-0.06</td><td>-4.18***</td></tr>
<tr><td>Repetition</td><td>-0.08</td><td>-0.11;-0.06</td><td>-6.78***</td></tr>
<tr><td>Cue Modality</td><td>-0.35</td><td>-0.40;-0.29</td><td>-11.60***</td></tr>
<tr><td>Semantic Rel.: Rep</td><td>0.02</td><td>0.004;0.035</td><td>2.43*</td></tr>
<tr><td>Semantic Rel.: Cue Mod.</td><td>0.05</td><td>0.02;0.09</td><td>3.05**</td></tr>
<tr><td>Repetition: Cue Mod.</td><td>0.05</td><td>0.03;0.06</td><td>5.80***</td></tr>
<tr><th colspan="4">Latency to First Fixation</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>0.77</td><td>0.68;0.85</td><td>17.84***</td></tr>
<tr><td>Semantic Relatedness</td><td>-0.10</td><td>-0.17;-0.04</td><td>-3.22**</td></tr>
<tr><td>Repetition</td><td>-0.02</td><td>-0.04;0.004</td><td>-1.63</td></tr>
<tr><td>Semantic Rel.:Repetition</td><td>0.04</td><td>0.02;0.06</td><td>3.26**</td></tr>
<tr><th colspan="4">Dwell Time</th></tr>
<tr><td>Predictor</td><td>β</td><td>CI
(2.5%; 97.5%)</td><td>t-value</td></tr>
<tr><td>Intercept</td><td>-2.94</td><td>-3.40;-2.49</td><td>-12.68***</td></tr>
</table>

(*) p < .10, *p < 0.05, **p < .01, ***p < .001

RTs were significantly faster for semantically unrelated critical objects compared to related ones when the cue was a picture [pairwise-comparison; *z-ratio* (3.75), p < .05]. Again, we observed faster RTs for increasing repetition, especially when participants were cued with a word. The latency to first fixation was also significantly faster for semantically unrelated critical targets (Figure 2B, right panel), but this effect was significantly reduced for increasing repetitions.

We did not observe any significant main effect or interaction on the dwell time (Figure 2c, right panel).

Discussion

The core objective of this study was to investigate the role that extrafoveal semantic processing plays in dual coding theory. Thus, we examined how the modality of stimuli (pictures vs. words) affects manual and eye movement responses during a visual search task. In line with the dual coding theory (Paivio, 1971; Paivio and Beggs, 1974), visual search for pictures was faster than for words. Along with the DCT, pictures activate a verbal and a visual representation, and this provides a pictorial advantage over words which instead only activate a verbal representation. Moreover, when cues must be converted (i.e., word-to-picture and picture-to-word) RTs were significantly slower compared to when the cue and the search array were in the same modality. This conversion cost was substantially lower when the cued word needed to be found in a picture array. Following the logic of the DCT, searching a picture into a word array should be faster than vice-versa, because the former case should benefit from dual coding compared to the latter, thereby facilitating the identification of the search target. This is not what we found. We speculate instead that what we observe relates to the nature of the search task. A cue word evokes a range of pictorial exemplars rather than a specific template, which speeds up the suppression of irrelevant pictorial distractors by simply relying on their verbal code. On the contrary, when searching for a picture in a word array, the activation of a perceptual template must be suppressed to find instead its orthographical representation. More interesting are the modulatory effects arising from semantic processing and the associated oculomotor search dynamics in extra-foveal and foveal vision. In line with previous research and our initial hypothesis, we confirm that semantic information is accessed even in the absence of perceptual template guidance (i.e., in target-absent trials). Independently of the modality of the cue, semantic information of visual objects (either as words or pictures) in the array is acquired in extrafoveal vision, used to prioritise early attention and drive overall faster response times when the critical object is semantically unrelated to the surrounding distractors (e.g., Cimminella et al., 2020; Telling et al., 2010 for picture arrays and Hohenstein and Kliegl, 2014 for word arrays). Critical to our study are those conditions that required conversion between codes. In such cases, we found that a semantically unrelated critical object speeds up search performance in both types of arrays, even if this effect in picture arrays and on the latency to first fixation is significant when the cue is a word. When looking at the foveal measure of dwell time, we observe longer dwell on semantically unrelated than related targets, corroborating previous research (e.g., Cimminella et al., 2020). However, this effect is present only in picture arrays possibly because words lack a distinctive perceptual representation beyond the orthographic form, unlike pictures, which instead have specific low-level visual features (e.g., shape or colour). Possibly, this result also points to a parallel search strategy in picture arrays as proposed by Hurley et al. (2021). When the target is semantically related to the distractors, all objects must be examined to verify the presence (or absence) of the target, while when it is unrelated to them, it creates a "pop-out" effect focusing foveal processing on it, so significantly reducing fixations onto the other objects. Finally, we also found an expected effect driven by the repetitions, which significantly improved search times (e.g., Malcolm & Henderson, 2009; Vickery et al., 2005) while reducing effects driven by the semantic unrelatedness of the critical object.

In sum, our study amply confirmed that semantic information is acquired in extra-foveal vision and it directly mediates the allocation of visual attention as well as overall search times. To this main finding, we add that semantic information also influences the effort to perform a code conversion (word-to-picture, picture-to-word). Concerning the DCT, we confirm that the picture-to-picture condition is the most efficient but, at the same time, we observe a smaller cost to convert a word cue (i.e., faster search times), which has a hypothetical single code, into a pictorial representation to be searched for. This finding may imply that it is the nature of the search task that may make the representational format of the cued target matter. A word evokes a generic picture template and activates a set of conceptual features to look for (e.g. if the cue is a *parrot*, *bird*, *feathers*, etc.). So, there is no actual need to have a precise perceptual template to be successful. On the contrary, a picture cue is unnecessarily rich if only a word must be found. As this perceptual information is irrelevant to the task, it must be suppressed, and hence the cost. At this point, the fixation-related potentials analysis, will allow us to establish the patterns of neural activity underlying the allocation of overt attention. This step will make it possible to identify the temporal interplay between semantic processing and representational access of words and pictures that is presumed by the DCT.

Acknowledgements

This work has been supported by NextGeneration EU under Piano di Ripresa e di Resilienza (PRIN 2022) with grant agreement (2022APAFFN; CUP B53D23014480001) awarded to MIC.

References

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading?. *Language and Linguistics Compass*, *13*(7), e12344.

Antúnez, M., López-Pérez, P. J., Dampuré, J., & Barber, H. A. (2022). Frequency-based foveal load modulates semantic parafoveal-on-foveal effects. Journal of Neurolinguistics, 63, 101071.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1).

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. Perception & Psychophysics, 70(8), 1444-1458.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PloS one*, *5*(5), e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PloS one*, *9*(9), e106953.

Cimminella, F., Sala, S. D., & Coco, M. I. (2020). Extra-foveal processing of object semantics guides early overt attention during visual search. *Attention, Perception, & Psychophysics*, *82*, 655-670.

Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related brain potentials during semantic integration of object–scene information. Journal of Cognitive Neuroscience, 32(4), 571-589.

Dampuré, J., Benraiss, A., & Vibert, N. (2019). Modulation of parafoveal word processing by cognitive load during modified visual search tasks. Quarterly Journal of Experimental Psychology, 72(7), 1805-1826.

Dampuré, J., Ros, C., Rouet, J. F., & Vibert, N. (2014). Task-dependent sensitisation of perceptual and semantic processing during visual search for words. *Journal of Cognitive Psychology*, *26*(5), 530–549.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166.

Hurley, R. S., Sander, J., Nemeth, K., Lapin, B. R., Huang, W., & Seckin, M. (2021). Differential eye movements in verbal and nonverbal search. Frontiers in Communication, 6, 654575.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. Perception, 36(14), 1-16

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13).

LaPointe, M. R., & Milliken, B. (2016). Semantically incongruent objects attract eye gaze when viewing scenes for change. *Visual Cognition*, *24*(1), 63-77.

Léger, L., Rouet, J. F., Ros, C., & Vibert, N. (2012). Orthographic versus semantic matching in visual search for words within lists. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 66(1), 32.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. Behavior research methods, 49, 1494-1502.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. Journal of Vision, 9(11), 8-8.

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. Journal of Vision, 10(2), 4-4.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305–315.

Nuthmann, A., De Groot, F., Huettig, F., & Olivers, C. N. (2019). Extrafoveal attentional capture by object semantics. PLoS One, 14(5), e0217051.

Paivio, A., & Begg, I. (1974). Pictures and words in visual search. Memory & Cognition, 2(3), 515-521.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. In *Quarterly Journal of Experimental Psychology* (Vol. 62, Issue 8).

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *40*(4), 473.

Rusich, D., Arduino, L. S., Mauti, M., Martelli, M., & Primativo, S. (2020). Evidence of semantic processing in parafoveal reading: A Rapid Parallel Visual Presentation (Rpvp) study. *Brain Sciences*, *11*(1), 28.

Telling, A. L., Kumar, S., Meyer, A. S., & Humphreys, G. W. (2010). Electrophysiological evidence of semantic interference in visual search. *Journal of Cognitive Neuroscience*, *22*(10), 2212-2225.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of vision*, *5*(1), 8-8.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 419.

Yan, M., Wang, A., Song, H., & Kliegl, R. (2019). Parafoveal processing of phonology and semantics during the reading of Korean sentences. *Cognition*, *193*, 104009.

Zelinsky, G. J. (2008). A Theory of Eye Movements During Target Acquisition. *Psychological Review*, *115*(4), 787–835