

Processos *bottom-up* e *top-down* no rastreamento ocular de imagens

Marcus Maia (UFRJ/CNPq) *

RESUMO: O artigo reporta experimento de monitoramento ocular de imagens em que 27 sujeitos foram expostos durante 10 segundos a versões de imagem que continham, respectivamente, elementos [-animado - humano], [+animado - humano] ou [+ animado + humano]. Cada versão podia ser precedida ou não de informação escrita presente na tela por 5 segundos, indicando o tema da imagem. Cada sujeito reportava, então, o que havia visto, registrando-se fixações e movimentos sacádicos (medidas *on-line*), bem como a reportagem final (medida *off-line*), concluindo-se que os tópicos previamente apresentados (efeito *top-down*), embora influenciem a medida *off-line*, não sobrepõem os traços estruturais salientes (*bottom-up*), na varredura *on-line*.

Palavras-chave: Psicolinguística; Rastreamento ocular de imagens; Efeitos top-down e bottom-up.

Introdução

Os estudos de rastreamento ocular (*eye-tracking*) tiveram seu início, como relatam Just & Carpenter (1976), com a pesquisa especificamente voltada para a leitura. Em 1878, o oftalmologista francês Louis Javal reporta estudo em que observou a olho nu crianças lendo textos, concluindo que seus olhos não pareciam mover-se continuamente ao longo das linhas, mas faziam uma série de pausas separadas entre si por pequenos saltos. Esses movimentos, que chegam a atingir velocidades altíssimas de até 700 graus por segundo (cf. CARPENTER, 1988), são denominados de sacadas, alternando-se com momentos de pausa, com duração média de 200-250 ms, denominados fixações. Muito embora as primeiras pesquisas tenham se orientado para a investigação dos movimentos sacádicos, mais do que para a análise das fixações, descobriu-se, posteriormente, o fenômeno conhecido como supressão sacádica (cf. IRWIN, 2004) em que o fluxo de informação visual é interrompido durante o movimento sacádico. Os estudos tendem, então, a concentrarem-se na duração das fixações, tomando-se, entretanto, a localização dos movimentos sacádicos progressivos ou regressivos como índice importante dos processos atencionais na leitura.

O estudo da percepção visual de cenas remonta, conforme revisto em Henderson & Ferreira (2004), ao trabalho de Buswell (1935), que apresenta resultados de experimentos em que, utilizando equipamento engenhoso, que registrava em filme fotográfico a luz refletida na córnea de sujeitos observando cenas através das pás de um ventilador, pôde determinar a direção e a duração das fixações. Buswell conclui que as fixações não eram distribuídas aleatoriamente nas 55 fotografias que apresentou a 200 sujeitos, mas tendiam a agrupar-se em regiões informativas das cenas, relacionando, pioneiramente, os movimentos e fixações oculares a processos atencionais. Além de monitorar o olhar de sujeitos vendo cenas livremente, Buswell também fez algumas manipulações assistemáticas do que ele chamou de “mental set” dos sujeitos ao olhar as fotografias, concluindo que as instruções experimentais ou a leitura de um parágrafo de texto previamente à apresentação das imagens poderia influenciar significativamente como as pessoas olhavam as imagens. O psicólogo russo Alfred

* maiamarcus@gmail.com

Yarbus também reporta, em livro de 1967, uma série de estudos de monitoramento ocular, em que estabeleceu, em concordância com Buswell (1935), que os padrões de inspeção visual de cenas dependem tanto de propriedades informativas do estímulo quanto das tarefas e objetivos prévios do observador.

A literatura sobre a atenção visual parece concordar, portanto, desde há muito tempo, que dois processos complementares atuam na inspeção ocular de cenas: um mecanismo *bottom-up*, guiado por propriedades intrínsecas do estímulo e um mecanismo *top-down*, relacionado a fatores tais como a disposição interpretativa aos objetivos prévios do observador. Não há controvérsia importante na literatura sobre a existência desta determinação dual *bottom-up/top-down* na cognição visual de imagens. A questão relevante na pesquisa contemporânea passa a ser, conforme apontam Henderson & Ferreira (2004), de um lado, a determinação específica das propriedades do estímulo, que constitui a abordagem conhecida como “saliency map”, em que se procura identificar, com precisão, que regiões do estímulo visual seriam proeminentes, explicitando suas diferenças em termos de dimensões tais como cor, intensidade, contraste, simetria, contorno, orientação etc. De outro lado, um desafio para a pesquisa tem sido o de estabelecer como os fatores *top-down* podem ser incorporados em um modelo completo de cognição visual. Trata-se de uma questão relacionada, entre outras coisas, ao curso temporal de atuação dos fatores *bottom-up* e *top-down* e de sua integração em uma arquitetura da cognição. Nesse sentido, Long & Olszewski (1999), Henderson et alii (1999) estudaram comparativamente o papel de fatores *bottom-up* (propriedades do estímulo) e de fatores *top-down* (contextualização prévia) na exploração de imagens, concluindo que as pistas *bottom-up* parecem ser mais influentes do que as pistas *top-down*. Em uma avaliação ampla da questão, Henderson & Ferreira (2004) questionam se um progresso substancial nesta área poderia ser feito, no entanto, caso não se estabeleça mais claramente, por exemplo, se nos movimentos sacádicos iniciais na exploração de uma cena já se pode observar efeitos de informação sobre a chamada “*global scene gist*” ou tópico global da cena, no sentido de que os primeiros movimentos sacádicos e fixações já seriam conduzidos de forma *top-down*.

O presente estudo procura investigar, exatamente, o curso temporal do acesso a informações *bottom-up* e *top-down* na percepção visual de cenas estáticas. A questão do curso temporal do acesso a diferentes informações cognitivas tem sido um tema recorrente na pesquisa que temos desenvolvido no Laboratório de Psicolinguística Experimental (LAPEX-UFRJ/CNPq) sobre a compreensão de frases em português. Em Maia, Faria, Buarque & Alcântara (2003) reportam-se resultados comparativos entre dados oriundos de questionários (*off-line*) e dados obtidos através de leitura auto-monitorada (*on-line*) de frases apresentando três diferentes tipos de ambiguidade estrutural, concluindo que os estudos *off-line* apresentaram maior sensibilidade a efeitos semânticos e pragmáticos, enquanto que os resultados dos estudos *on-line* parecem sugerir que estes fatores não estariam imediatamente disponíveis ao processador sintático, diferentemente das informações estritamente estruturais que seriam acessadas rapidamente. Também em Maia, Fernández, Costa & Lourenço-Gomes (2007) comparamos dados *on-line* e *off-line* na leitura auto-monitorada de frases contendo orações relativas apostas a SN complexos, concluindo que a informação relativa ao comprimento da oração relativa é atuante apenas na medida *off-line*, não influenciando a medida *on-line*. No estudo que reportamos a seguir esta questão sobre o papel de fatores estruturais vis à vis o papel de fatores mais gerais, fundamental para o desenvolvimento de

modelos de arquitetura cognitiva, é investigada no que se refere à percepção visual de imagens estáticas, utilizando a técnica de monitoramento ocular.

1. O experimento de rastreamento ocular de imagens

O processamento visual de cenas é delimitado pelas propriedades neuro-anatômicas do sistema visual humano. A região que permite maior acuidade visual é a chamada região foveal, uma área do ponto de fixação, compreendendo entre um e dois graus do ângulo visual. A fóvea está centrada no eixo ótico do olho, tendo uma alta densidade de células especializadas, denominadas cones, foto-receptores responsáveis pela percepção visual de detalhes de forma e cor. A informação aí recebida é enviada para o córtex visual no cérebro, que destina grande parte de sua área para o processamento da informação foveal (magnificação cortical). A figura 1, abaixo, ilustra a região foveal, bem como as áreas de fixação foveal, para-foveal e periférica.

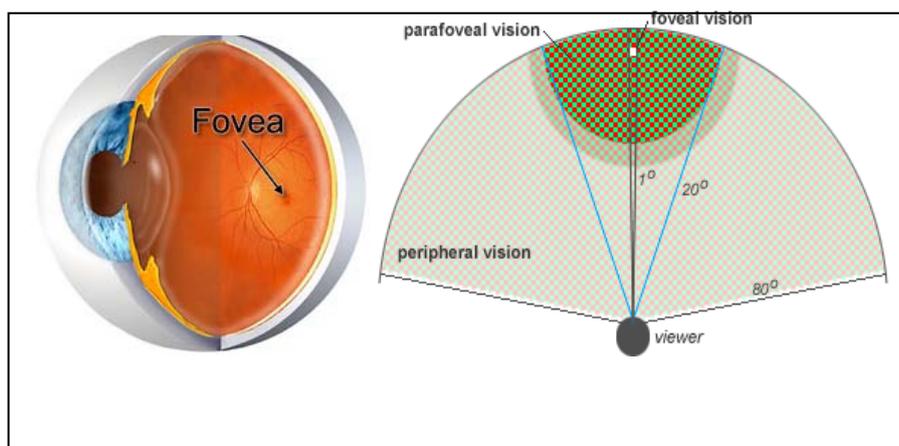


Figura 1 – Fóvea e campo visual

Para aproveitamento pleno do input foveal de alta qualidade, nas tarefas visuo-cognitivas complexas, como a leitura e a visualização de cenas, é necessário que a fixação foveal se desloque constantemente, em média três vezes por segundo, embora haja grande variabilidade de índices de fixação na visualização de imagens, como reporta Henderson (2003). Além dos pontos de fixação foveal, outro parâmetro importante para a análise dos padrões de visualização de cenas é a direção dos movimentos sacádicos. Ao contrário do que se tem observado na leitura, na visualização de imagens não se pode estabelecer uma assimetria da fixação foveal e nem no padrão de movimentação sacádica. Na leitura, cada fixação captura cerca de 2/3 de informação à direita e 1/3 à esquerda, nos leitores de línguas com escrita da esquerda para a direita (RAYNER, WELL & POLLATSEK, 1980). Em leitores de hebraico, que se escreve da direita para a esquerda, a assimetria é revertida (POLLATSEK, BOLOZKY, WELL & RAYNER, 1981). Na exploração visual de imagens tal assimetria não é observada. Como revisamos na introdução deste artigo, a varredura de

cenar é determinada de forma dual por propriedades do estímulo e também por fatores do tipo *top-down*. Tarefas de categorização imediata de cenas têm estabelecido um limiar de 20 ms de exposição para a categorização de cenas em termos de hiperônimos com índice de acerto maior do que 90% e tempos de decisão médios abaixo de 400 ms (VAN RULLEN & THORPE, 2001), indicando que as primeiras fixações já permitem uma caracterização básica da cena.

O presente experimento tem como principal objetivo verificar a interação entre propriedades intrínsecas (fatores *bottom-up*) de uma imagem e a existência ou não de um tópico apresentado previamente à exibição da imagem (fatores *top-down*). De modo mais específico, examinamos se a existência de um título anterior à visualização da imagem é capaz de influenciar a sua exploração visual através do padrão de movimentos sacádicos e de fixações (medida *on-line*) e de reportagem final sobre o conteúdo da imagem, após a sua visualização (medida *off-line*), procurando, portanto discriminar o curso cronológico dos fatores *bottom-up* e *top-down*. As variáveis independentes do experimento são:

- a) Propriedades do estímulo (casal, cão, nada)
- b) Título prévio (casa, pedras, nada)

O *design* experimental cruzava os três níveis do fator (a) com os três níveis do fator (b), produzindo nove condições experimentais, cada uma das quais testada em três sujeitos, totalizando 27 sujeitos (9x3). A imagem utilizada para apresentação aos sujeitos por lapso de tempo pré-fixado em dez segundos foi o quadro “Idílio” de Tarsila do Amaral, tendo-se manipulado a imagem com o programa *Photoshop*, para gerar os três níveis da variável independente (a), ilustrados abaixo:



Figura 2 - Propriedades do estímulo: casal [+animado,+humano]



Figura 3 – Propriedades do estímulo: cão [+animado, -humano]



Figura 4 – Propriedades do estímulo: nada [-animado, -humano]

A variável independente (b), título prévio, consistia na apresentação por cinco segundos, anteriormente à exibição da imagem, de uma de três telas, contendo o título “Casa na colina”, “Pedras no Riacho” ou nenhum título (tela em branco).

A tarefa experimental consistia na observação da imagem seguida da redação de um parágrafo, reportando o conteúdo da imagem observada. As variáveis dependentes foram três, a saber, movimentos sacádicos iniciais, tempos totais de fixação, reportagem final. Para efeitos de análise, estabeleceram-se três regiões de interesse para as medidas *on-line*, onde foram computados os movimentos sacádicos iniciais, bem como os tempos totais de fixação, conforme ilustrado abaixo:



Figura 5 – Regiões de interesse para as medidas *on-line*

Como se pode ver na Figura 5, as regiões de interesse eram a região onde se manipulou a variável independente (a) “propriedades do estímulo” (casal, cão, nada) e as duas regiões correspondentes a dois níveis da variável independente (b) “título prévio” (casa na colina, pedras no riacho). As duas medidas *on-line*, a saber, os índices de movimento sacádicos iniciais e os tempos totais de fixação foram computados para cada uma dessas regiões.

1.1. Método

Participantes

Participaram do experimento, como voluntários, 27 alunos de graduação, cursando a disciplina Psicolinguística, no quarto período do curso de Fonoaudiologia da UFRJ, no primeiro semestre de 2008, todos do sexo feminino, com idade média de 20 anos, com visão normal, sem uso de óculos ou lente de contato.

Material

Manipulou-se com o programa *Photoshop* a imagem “Idílio” de Tarsila do Amaral, produzindo-se as três cenas correspondentes à variável independente (a), propriedade do estímulo, ilustradas nas figuras 2, 3 e 4, que eram precedidas por uma das telas correspondentes aos níveis da variável independente (b), título prévio, a saber, “pedras no riacho”, “casa na colina” ou tela em branco. As telas relativas à variável (a) permaneciam por 10 segundos para observação dos sujeitos e eram precedidas pelas telas relativas à variável (b) que permaneciam por 5 segundos para observação. Cada uma das 9 condições experimentais geradas pelo cruzamento das duas variáveis independentes foi testada em três sujeitos, totalizando 27 sujeitos.

O aparato experimental consistiu no equipamento Arrington View Point Quick Clamp Eye-Tracker, um sistema que monitora o olho direito, com base em informações obtidas através do movimento da pupila e do reflexo da córnea, com resolução temporal de 30 Hz. As telas foram apresentadas em monitor de 17 polegadas, que se encontrava a distância de 60 cm dos olhos dos sujeitos. Observe-se que o programa que controla o rastreador é acessado em monitor de 17 polegadas, distinto do monitor em que os estímulos experimentais são

apresentados. Utilizou-se um descanso de queixo e um clipe nasal para minimizar os movimentos de cabeça durante a visualização, conforme ilustrado na Figura 6:

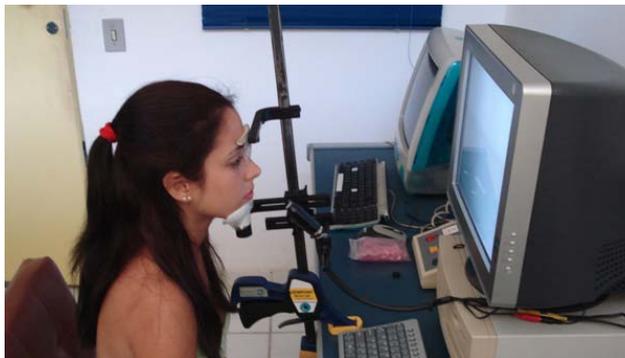


Figura 6 – Aparato experimental

Procedimento

Os participantes foram testados individualmente na sala do LAPEX, na qual o aparato encontra-se instalado. Inicialmente, o experimentador orientava cada participante, esclarecendo que a tarefa consistia na visualização de duas telas, a primeira das quais permaneceria por 5 segundos no monitor, enquanto que a segunda permaneceria por dez segundos. Solicitava-se que o participante observasse atentamente a imagem na segunda tela, de modo que, logo após o encerramento do experimento, que era assinalado por uma tela com a palavra FIM, o participante utilizasse a folha e a caneta na mesa ao lado para redigir um parágrafo descrevendo a imagem que acabara de ver.

Em seguida, o sujeito era acomodado na estrutura do rastreador, ajustando-se o aparato de acordo com a sua altura e características físicas. Fazia-se, nesse momento, o ajuste denominado de “threshold”, utilizando-se o monitor de controle, enquanto o sujeito visualizava um ponto fixo na tela do seu monitor, garantindo-se que o foco da micro-câmera tenha capturado a pupila do sujeito, conforme indicado na Figura 7, que ilustra a imagem do monitor de controle enquanto o sujeito fixa seu olhar em um ponto central do seu monitor. Ao final desta etapa solicitava-se ao participante que fixasse seu olhar nos quatro cantos da tela do seu monitor, alternadamente, observando-se, no monitor de controle, que a sua pupila continuava envolta pelo círculo amarelo, indicando sua captura.

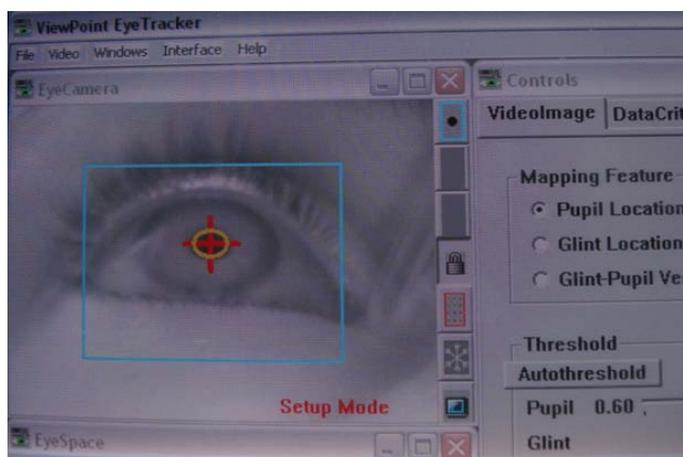


Figura 7 – Ilustração do processo de *threshold*

Tinha, então, início o processo conhecido como calibração, durante o qual o programa do rastreador era ajustado às características de movimentação ocular de cada sujeito. A calibração consiste na observação de pontos que surgem aleatoriamente em diferentes regiões da tela e que devem ser fixados pelo sujeito até que desapareçam. Imediatamente após o desaparecimento de um ponto, um outro ponto surge na tela, devendo ser imediatamente observado pelo sujeito que deve, novamente, manter seu olhar fixo até que o ponto desapareça. Após a calibração, o experimentador indicava que o experimento iria, de fato, começar. Após a apresentação das duas telas experimentais (5 segundos para a primeira tela e 10 segundos para a segunda tela, que continha a imagem), uma tela com a palavra FIM era apresentada, devendo, então o sujeito afastar-se do equipamento e redigir o parágrafo de reportagem do conteúdo da imagem, conforme a instrução prévia. De modo geral, a duração total de cada sessão foi de aproximadamente 10 minutos.

1.2. Resultados e discussão

1.2.1. Figura sem elementos [+humano] ou [+animado]

Apresenta-se inicialmente o conjunto de resultados para as medidas relativas à Figura 4, que não continha nem o casal e nem o cachorro. De modo geral, esta imagem apresenta os maiores índices de interferência dos fatores *top-down*, tanto nas medidas *on-line*, quanto nas medidas *off-line*, em relação às imagens com o casal e com o cão.

Figura sem elementos [+humano] ou [+animado]

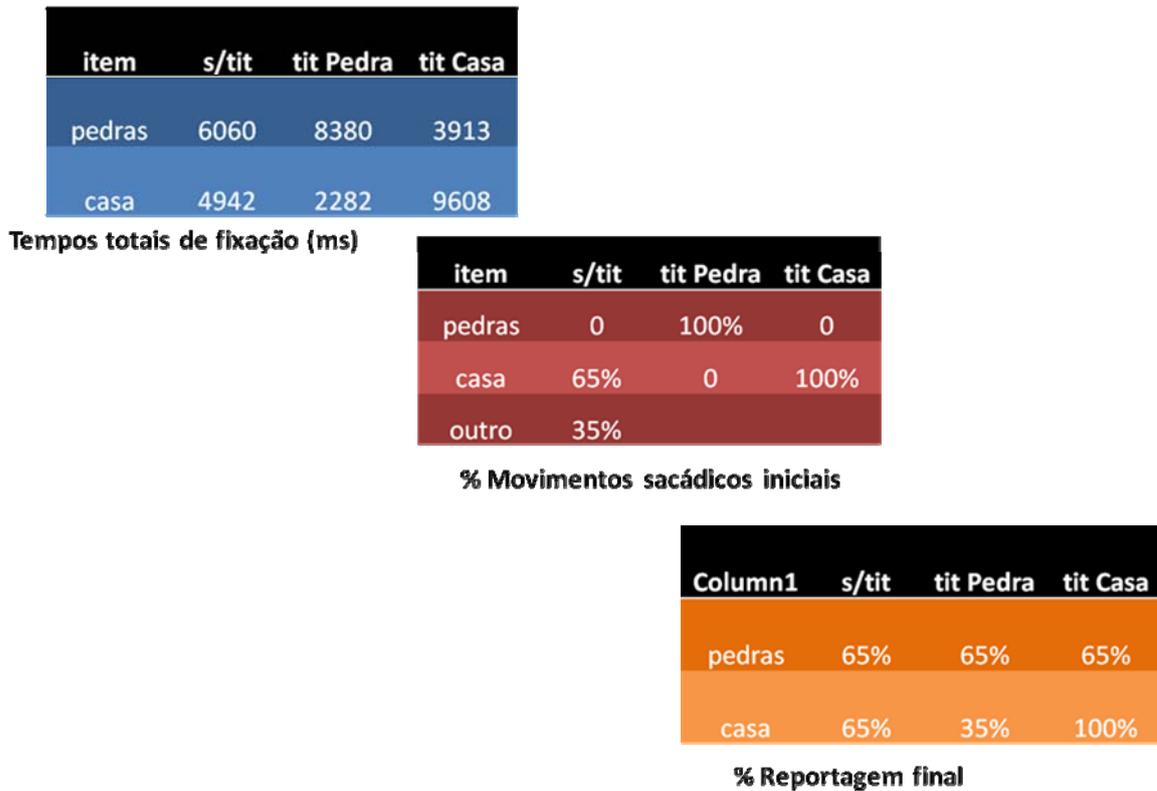


Figura 8- Resultados das medidas *on-line* e *off-line* relativas à figura sem o casal e sem o cão

1.2.1.1. Tempos totais de fixação

Observe-se que, quando não estão presentes na imagem o casal ou o cão, os tempos totais de fixação nas regiões de interesse relativas à casa na colina e às pedras no riacho não variam significativamente entre si, na condição em que não se fornece título na primeira tela ($t=1,126$, $p= 0,27$). Por outro lado, na condição em que o título é “Pedras no riacho” os tempos de fixação na região de interesse relativa às pedras é significativamente maior do que os tempos de fixação na região da casa na colina ($t=2,161$, $p=0,038$). De modo equivalente, também há diferença significativa entre os tempos totais de fixação na região das pedras e os tempos totais de fixação na região da casa na colina, na condição em que o título é “A casa na colina”. Crucialmente a diferença se dá na direção esperada, ou seja, o título determina maiores tempos de fixação na região da casa do que na região das pedras, muito embora a área física da região das pedras seja maior do que a área de interesse em que se encontra a casa ($t=4,094$, $p=0,0002$).

1.2.1.2. Percentual de movimentos sacádicos iniciais

A medida relativa aos índices de movimentos sacádicos iniciais é consistente com os achados reportados na seção acima, relativos aos tempos de fixação totais durante a visualização da imagem. Quando não estão presentes na imagem nem o casal e nem o cão, as pedras nunca são alvo das primeiras sacadas quando não se fornece título para a imagem, enquanto que a casa passa a ser objeto de 65% das sacadas iniciais. Quando o título faz alusão às pedras, os primeiros olhares são sempre direcionados à região das pedras, na imagem. De forma equivalente, quando o título se refere à casa, os movimentos sacádicos iniciais são sempre dirigidos à região da casa. A figura 9, abaixo, ilustra um desses casos, em que após ler o título “A casa na colina” o sujeito visualiza a imagem, fazendo uma primeira fixação central e um primeiro movimento sacádico na direção da casa (traço em vermelho), seguido de fixação nesta região (mancha verde).

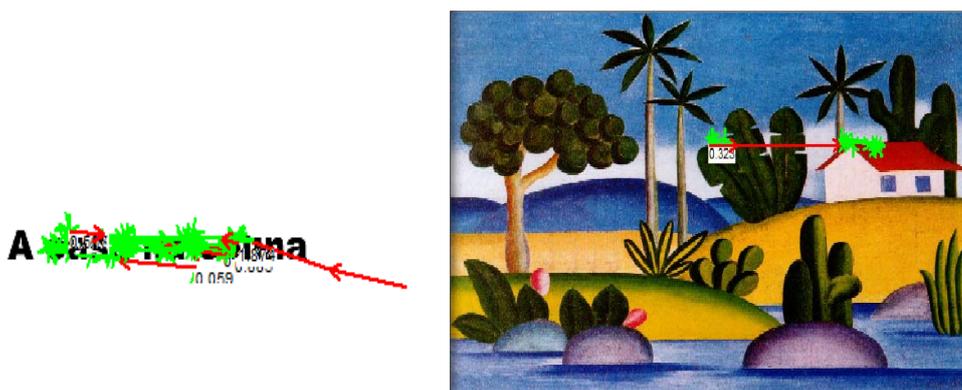


Figura 9 – Rastreamento ocular na condição com título “Casa na colina” e imagem sem o casal e sem o cão.

1.2.1.3. Percentual de referências na reportagem final

Medem-se, aqui, os índices percentuais de referências feitas à casa e às pedras no riacho na reportagem final (medida *off-line*), solicitada logo após a visualização da imagem sem o casal e sem o cachorro. Observe-se que, na condição sem título, a casa na colina e as pedras no riacho apresentam os mesmos percentuais de referência no parágrafo de reportagem final escrito pelos sujeitos. Por outro lado, quando o título é “Pedras no riacho”, as pedras são mencionadas em índices superiores (65%) à menção feita à casa (35%). Quando o título é “A casa na colina”, a menção à casa atinge o nível de 100% nas reportagens, enquanto as pedras ficam em 65%.

1.2.2. Figura com o elemento cão [-humano] e [+animado]

Apresentamos abaixo as tabelas com os resultados obtidos nas três medidas quando a figura apresentava o cão. Em comparação com as medidas reportadas para a condição em que a figura não exibia nem o cão e nem o humano, que permitiram estabelecer a existência de efeito *top-down* resultante da ação do título prévio à visualização da imagem, pode-se observar, agora, que a introdução do elemento cão já dificulta que este efeito se instancie de modo tão claro. De modo geral, nas três medidas, observe-se que, agora, apenas o título “Casa na colina”, relacionado a traço [+humano], apresenta efeitos significativos. O título “Pedras no riacho”, por outro lado, tem seu efeito reduzido em função da presença do cão na imagem.

Figura com elemento [+animado]

Column1	s/tit	tit Pedra	tit Casa
cão	5971	5614	2208
pedras	5306	6085	6934
casa	5579	5213	8405

Tempos totais de fixação (ms)

Column1	s/tit	tit Pedra	tit Casa
cão	1/3	1/3	0
pedras	0	2/3	0
casa	1/3	0	3/3
outro	1/3		

% Movimentos sacádicos iniciais

Column1	s/tit	tit Pedra	tit Casa
cão	100%	100%	65%
pedras	65%	100%	65%
casa	100%	100%	100%

% Reportagem final

Figura 10- Resultados das medidas on-line e off-line relativas à figura com o cão

1.2.2.1. Tempos totais de fixação

Os tempos totais de fixação nas três regiões de interesse não diferem significativamente entre si quando a imagem contendo o cão é exibida sem apresentação de

qualquer título prévio. A comparação dois a dois revela falta de significância tanto quando se comparam os tempos totais de fixação entre o cão e as pedras ($t=0,06$, $p=0,95$), quanto entre o cão e a casa ($t=0,45$, $p=0,65$) e entre as pedras e a casa ($t=0,38$, $p=0,70$). Na condição em que o título é “Pedras no riacho”, ao contrário do que se observou na seção 1.2.1.1, em que há efeito deste título sobre os tempos de fixação na imagem que não apresenta nem o cão e nem o humano, agora já não há efeito *top-down* significativo deste título sobre os tempos de fixação nas três regiões de interesse. O cão e as pedras recebem fixações que não permitem rejeitar a hipótese nula ($t=0,79$, $p=0,43$). Igualmente não há significância na diferença de tempos totais de fixação entre o cão e a casa ($t=0,36$, $p=0,71$) ou entre as pedras e a casa ($t=0,42$, $p=0,67$). Quando o título é “Casa na colina”, também não se obtém significância na comparação entre os tempos totais de fixação recebidos pelo cão e as pedras ($t=0,12$, $p=0,89$) ou quando se compara as pedras e a casa ($t=1,58$, $p=0,12$). Significância só é obtida quando se comparam os tempos do cão com a casa, onde o título efetivamente impõe maiores tempos de fixação na região da casa ($t=2,68$, $p=0,01$).

1.2.2.2. Percentual de movimentos sacádicos iniciais

Nesta medida, pode-se observar que, de modo semelhante, na imagem em que não havia a presença do cão ou do casal, analisada em 1.2.1.2, a região das pedras no riacho continua a não atrair os primeiros movimentos sacádicos na condição em que não se apresenta título prévio. Entretanto, o mapa de elementos salientes é alterado quando o cão entra na figura, no que se refere à competição com a casa. Observe ainda nesta condição em que não se apresenta título prévio que o cão e a casa recebem os mesmos índices de sacadas iniciais, ao contrário do que foi registrado em 1.2.1.2, em que a casa na figura sem o cão ou humano recebia a maior parte das sacadas iniciais na condição sem título prévio. Quando se apresenta previamente o título “Pedras no riacho”, as pedras recebem 2/3 das sacadas iniciais e o cão fica com 1/3 desses movimentos. Entretanto, note-se que o cão impõe uma perda para as pedras, comparativamente ao registrado para a figura sem o cão ou sem o humano, analisada em 1.2.1.2, em que as pedras receberam 100% das sacadas iniciais quando o título era “Pedras no riacho”. Por outro lado, há um efeito *top-down* claro quando o título é “A casa na colina”, pois nesta condição a casa recebe 100% das sacadas iniciais a despeito da presença do cão, o que já sugere a relevância de elemento associado ao traço [+humano] no mapa de elementos salientes no *input*, pois o título “Pedras no riacho” por si só, como vimos acima, não garante a totalidade dos movimentos sacádicos iniciais.

1.2.2.3. Percentual de referências na reportagem final

Na condição sem título prévio, o cão e a casa são sempre mencionados, enquanto as pedras deixam de ser referidas em 35% das reportagens finais. Na condição em que o título é “Pedras no riacho”, as três regiões de interesse (pedras, casa e cão) recebem sempre menção em todas as reportagens. Na condição em que o título é “A casa na colina”, a casa é mencionada em 100% das reportagens, enquanto que o cão e as pedras ficam com 65% das menções, cada um.

1.2.3. Figura com elemento casal [+humano] e [+animado]

Nesta seção apresentam-se os resultados das três medidas quando a imagem exibida inclui o casal. Nessa condição fica patente o efeito do traço [+humano] no mapa de elementos salientes do input em todas as medidas. O desbalanceamento do efeito *top-down* observado na condição da imagem sem o casal e sem o cão reportado na seção 1.2.1, que já pôde ser notado na condição da imagem com o cão (seção 1.2.2), é agora claramente instanciado na condição da imagem com o casal. Nessa condição, o efeito *bottom-up* do input é patente nas três medidas, com prejuízo significativo para o efeito *top-down*, ao menos nas duas medidas *on-line*.

Figura com elemento [+humano]

Column1	s/tit	tit Pedra	tit Casa
casal	11144	10308	7213
pedras	3936	7287	6683
casa	3197	6299	5254

Tempos totais de fixação (ms)

Column1	s/tit	tit Pedra	tit Casa
casal	100%	100%	100%
pedras	0	0	0
casa	0	0	0

% Movimentos sacádicos iniciais

Column1	s/tit	tit Pedra	tit Casa
casal	100%	100%	100%
pedras	0	100%	0
casa	65%	65%	100%

% Reportagem final

Figura 11- Resultados das medidas *on-line* e *off-line* relativas à figura com o casal

1.2.3.1. Tempos totais de fixação

Na condição sem título, os tempos totais de fixação na região do casal são significativamente maiores do que os tempos observados na região das pedras ($t = 2,61$,

$p=0,01$) e na região da casa ($t= 2,8$, $p=0,008$). Os tempos de fixação não diferem quando se comparam as pedras com a casa ($t=0,17$, $p=0,87$). Na condição em que o título é “Pedras no riacho”, as pedras já não recebem fixações significativamente maiores nem em relação ao casal ($t=1,44$, $p=0,16$) e nem em relação à casa ($t=1,23$, $p=0,23$), diferentemente do observado na condição em que a imagem não exibia nem o casal e nem o cão (1.2.1.1), em que os tempos de fixação na região de interesse relativa às pedras foram significativamente maiores do que os tempos de fixação na região da casa na colina, por força do título “Pedras no riacho”. Agora, o efeito *top-down* do título é completamente anulado pelo efeito *bottom-up* representado pela presença do casal no mapa de elementos salientes do *input*. Também quando o título é “A casa na colina” o efeito *top-down* do título não se instancia nem em relação ao casal ($t=0,75$, $p=0,45$) e nem em relação às pedras ($t=0,064$, $p=0,99$). Note-se que o efeito *top-down* do título “Pedras no riacho” não ocorre mesmo em uma comparação com os tempos de fixação na região das pedras na condição sem título ($t=2,007$, $p=0,054$).

1.2.3.2. Percentual de movimentos sacádicos iniciais

No que se refere aos movimentos sacádicos iniciais, a presença do casal no *input* atrai sempre de forma consistente as primeiras sacadas, que nunca são dirigidas para as demais regiões de interesse, quer na condição sem título, quer nas condições em que os títulos fazem referência às pedras ou à casa. O padrão de movimento sacádico inicial em direção ao casal, ilustrado na figura 12, comprova nesta medida o predomínio do efeito *bottom-up* sobre o *top-down*.

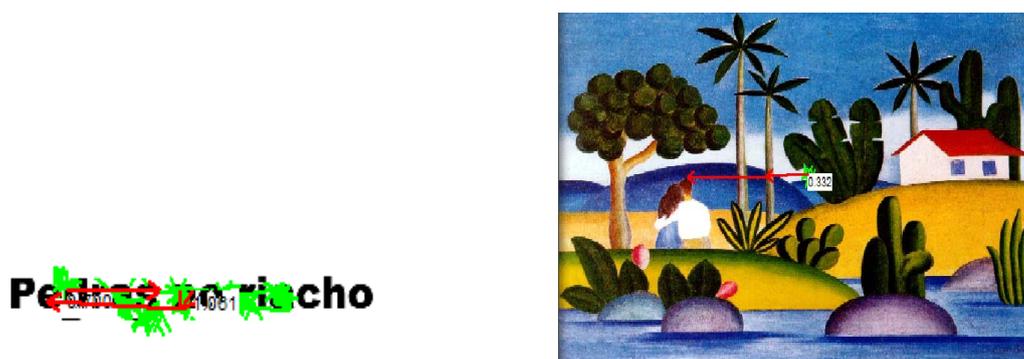


Figura 12 – Rastreamento na condição com título “Pedras no riacho” e imagem com o casal.

É interessante ainda observar que a presença do casal no *input* é um atrator *bottom-up* tão influente que, mesmo em um controle realizado em que se inverteu a figura do casal para o lado direito da imagem, este elemento continuou a receber a sacada inicial no escaneamento visual da imagem.

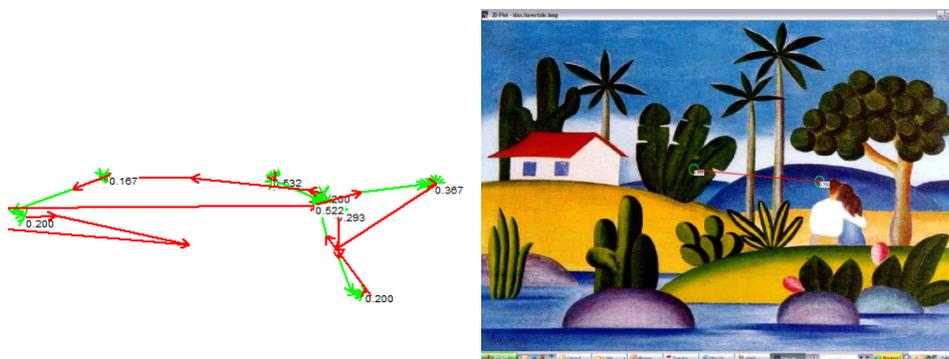


Figura 13 – Rastreamento em condição sem título com casal á direita da imagem

1.2.3.3. Percentual de referências na reportagem final

Nesta condição, a referência ao casal é feita em 100% das reportagens finais, independentemente de haver título ou não, comprovando o efeito *bottom-up* já identificado nas medidas *on-line*. Registre-se, no entanto, que ao contrário das medidas *on-line*, esta medida indica efeito *top-down* do título – as pedras são mencionadas em 100% das reportagens que as mencionam como título e a casa, igualmente, é sempre mencionada quando o título a coloca em destaque.

Conclusões

Com base nos resultados apresentados e discutidos acima, conclui-se que:

1. A exploração visual das propriedades físicas de uma imagem (computação *bottom-up*) não é randômica. Há padrões de movimentos sacádicos e de latências de fixações relacionados a propriedades informativas ou salientes do *input*.
2. Não parece haver uma relação determinística entre a informação *top-down* e o escaneamento inicial da imagem. Os fatores *top-down* influenciam o olhar, mas podem ser sobrepujados por características intrínsecas salientes da imagem.
3. Efeitos *top-down* se instanciam mais claramente em medidas *off-line* do que em medidas *on-line*. Por outro lado, a computação *bottom-up*, capturada em medidas *on-line*, pode não se instanciar inequivocamente em medidas *off-line*.
4. O desenvolvimento de arquiteturas cognitivas que modelem a integração de algoritmos *bottom-up* e heurísticas *top-down* não pode prescindir, portanto, de pesquisa experimental sensível ao curso cronológico das operações atuantes no processo.

ABSTRACT: This article reports an eye-tracking experiment in which 27 subjects were exposed during 10 seconds to versions of an image which contained either a [- animate – human], a [+animate - human] or a [+animate +human] element. Each version could be preceded or not by written information which remained on the screen during 5 seconds, indicating the topic of the image. Fixation times and saccadic movements (on-line measures) were registered and subjects were also asked to write a one-paragraph report (off-line measure) immediately after viewing the scene. Based on the results obtained it is suggested that the previously presented topics (top-down effect) may influence the off-line measure, but cannot override the bottom-up computation of salient elements in the input in the on-line measures.

Key-words: Psycholinguistics; Eye-movements in scene viewing; Top-down and bottom-up effects.

Referências

- BUSWELL, G. T. **How People look at pictures**. Chicago: University of Chicago Press, 1935.
- CARPENTER, R. H. S. **Movements of the Eyes**. London: Plon, 1988.
- HENDERSON, J. M. Human gaze control in real-world scene perception. **Trends in Cognitive Sciences**, 7, 498-504, 2003.
- HENDERSON, J. M.; FERREIRA, F. (Eds.). **The interface of language, vision, and action: Eye movements and the visual world**. (pp 1-58). New York: Psychology Press, 2004.
- HENDERSON, J. M.; HOLLINGWORTH, A. The role of fixation position in detecting scene changes across saccades. **Psychological Science**, 10, 438-443, 1999.
- IRWIN, D. E. Fixation Location and Fixation Duration as Indices of Cognitive Processing. In HENDERSON, J. M.; FERREIRA, F. (Eds.). **The Integration of Language, Vision, and Action: Eye Movements and the Visual World**. New York: Psychology Press, 2004.
- JAVAL, L. E. Essai sur la physiologie de la lecture. **Annales d'Oculistique**, 1878, 82, 242-253.
- JUST, M. A.; CARPENTER, P. A. Eye fixations and cognitive process. **Cognitive Psychology**, 8, 441-480, 1976.
- LONG, G. M.; OLSZWESKI, A. D. To reverse or not to reverse: when is an ambiguous figure not ambiguous?. **American Journal of Psychology**. 112, pp. 41–71, 1999.
- MAIA, M. A. R.; FARIA, F. S.; BUARQUE, S.; ALCANTARA, S. N. O Processamento de concatenações sintáticas em três tipos de estruturas frasais ambíguas em português. **Forum linguístico**, Santa Catarina, v. 3, n. 1, p. 13-53, 2003.
- MAIA, M. A. R.; FERNÁNDEZ, E.; COSTA, A.; LOURENÇO-GOMES, M. do C. Early and late preferences in relative clause attachment in Portuguese and Spanish. **Journal of Portuguese Linguistics**, v. 5/6, p. 227-250, 2007.
- POLLATSEK, A.; BOLOZKY, S.; WELL, A. D.; RAYNER, K. Asymmetries in the perceptual span for Israeli readers. **Brain and Language**, 14, 174-180, 1981.
- RAYNER, K.; WELL, A. D.; POLLATSEK, A. Asymmetry of the effective visual field in reading. **Perception & Psychophysics**, 27, 537-544, 1980.
- VAN RULLEN, R.; THORPE, S. J. Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artificial objects. **Perception**, 30: 655-68, 2001.
- YARBUS, A. L. **Eye movements and Vision**. New York: Plenum Press, 1967.