2
Percepção visual e processamento da imagem

2.1
A percepção visual: aspectos biológicos

Os avanços nas Neurociências nas últimas décadas têm permitido uma explicitação mais precisa dos processos de percepção visual. Inicialmente, pode-se considerar a trajetória da luz no olho, que atravessa a córnea em direção à íris. A dilatação ou compressão da pupila permite que mais ou menos luz entre no olho. Eysenck & Keane (2007, p. 41) chamam esse processo de recepção do estímulo visual. Posteriormente, ocorre a transdução, em que a energia luminosa é convertida em padrões eletroquímicos nos neurônios. No processo de codificação, esses padrões eletroquímicos são interpretados pela atividade do sistema nervoso no cérebro.

Ainda no processo de recepção, deve-se destacar a importância do cristalino, que direciona a luz para a retina, localizada na parte posterior do olho. A retina, por sua vez, é um complexo tecido composto por cinco tipos de células diversas. A luz atravessa essas camadas de células, chega às receptoras e retorna por meio dessas camadas. No receptor óptico da retina existem dois tipos de células, os cones e os bastonetes. Os cones são mais especializados em cores e agudeza da visão (Eysenck & Keane, 2007, p. 42). Nas áreas externas da retina estão localizados cerca de 125 milhões de prismas, responsáveis pela percepção do movimento e pela visão no escuro. Finalmente, as informações visuais, sob a forma de impulsos nervosos, são dirigidas ao nervo óptico e seguem o caminho da retina-geníclulo-estriada, que, por sua vez, transmite a informação da retina para o córtex visual primário. No sistema retina-geníclulo-estriado, há dois canais, a via parvocelular (P), caracterizada por apresentar maior sensibilidade à cor e pequenos detalhes e por lidar com dados vindos dos cones; a via magnocelular (M) é caracterizada por maior sensibilidade ao movimento e lida com dados vindos dos bastonetes. As vias P e M não são totalmente isoladas. Segundo Nealey & Manusell (1994, *apud* Eysenck & Keane, 2007, p. 42), "há uma entrada de informações do caminho M para o caminho P".

O processamento visual no córtex visual primário (V1) e no secundário (V2) ocorre com base na noção de campo receptivo, que é a região da retina atingida pela luz. Além disso, o fenômeno da inibição lateral, isto é, a diminuição da atividade neuronal em função de outro neurônio ativado na vizinhança, é fundamental para aumentar o contraste das extremidades dos objetos. Isso permite que o observador identifique limites entre objetos.

Eysenck & Keane (2007) ressaltam que o córtex pode ser subdividido em mais de 30 áreas, das quais 25% delas são especializadas em visão. Por outro lado, muito do que já se descobriu sobre o córtex visual origina-se de estudos realizados com primatas não-humanos, e não se deve assumir que haja uma grande similaridade entre os sistemas visuais humanos e dos primatas não-humanos. Zeki (1992, 1993 *apud* Eysenck & Keane, 2007, p. 45), relacionou algumas áreas do córtex visual a certas funções. Por exemplo, V1 e V2 seriam ativadas no início da percepção visual, por apresentarem células que interagem com cor e forma e compilação de dados que são repassados a outras áreas. V3 e V3A apresentam maior especialização quanto à forma dos objetos, mas não quanto à cor; em V4, ocorre mais interação com a cor e orientação de linhas. V5 está mais especializada em movimento. As evidências que levaram os cientistas a essas conclusões foram obtidas em estudos com macacos. As evidências com seres humanos foram obtidas, de modo geral, em estudos com humanos que apresentavam lesões em diferentes áreas do cérebro. Por exemplo, pode-se confirmar a associação de V4 ao processamento de cor a partir de investigações com pacientes de acromatopsia, que é um impedimento no processamento de cor. Entretanto, os pacientes de acromatopsia também apresentam dano nas áreas V2 e V3. É importante ressaltar que V4 não é a única área responsável pelo processamento da cor. A figura abaixo destaca a localização das áreas do córtex cerebral acima apresentadas, conforme Fulton (2000):

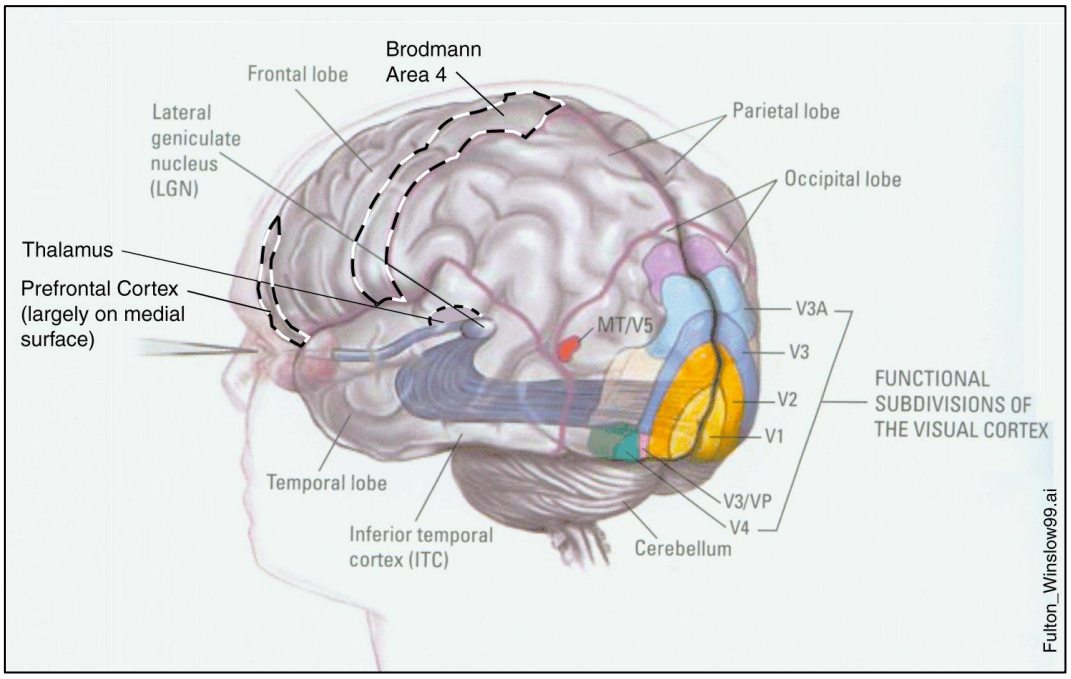

Figura 1 – Esquema com algumas áreas responsáveis pelo processamento visual no córtex cerebral (Fulton, 2000).

O reconhecimento de objetos é uma questão primordial na descrição da percepção visual. Os psicólogos da Gestalt tentaram analisar esse processo, tendo desenvolvido a lei da Prägnanz: "Dentre várias organizações possíveis, será percebida aquela que se caracterizar pela melhor forma, a mais simples e estável" (Koffka, 1935 *apud* Eysenck, Keane, 2007, p. 75). A Gestalt surgiu como reação ao enfoque mais estruturalista da percepção, segundo o qual a percepção da forma é feita pelo desmembramento do todo em componentes elementares. A Gestalt propunha uma orientação mais funcional, tencionando tratar de processos globais e holísticos envolvidos na percepção da estrutura do ambiente que, por sua vez, seria influenciada pelo conhecimento prévio ou enciclopédico do observador.

Outro aspecto enfatizado pelos gestaltistas foi a distinção entre figura e fundo na organização perceptual – basicamente, a figura (objeto em foco) teria uma forma mais definida, em oposição ao fundo, de forma menos definida. Contudo, há problemas na abordagem da Gestalt. Por exemplo, a maior parte das evidências obtidas vinha de desenhos lineares e bidimensionais. Além disso, muitos dados tinham um caráter meramente descritivo, sem que houvesse explicações mais substanciais dos fenômenos.

A influência na percepção visual de fatores mais diretamente associados a propriedades dos próprios estímulos – chamados fatores *bottom-up* (de baixo para cima) e do conhecimento prévio do observador – fatores *top-down* (de cima para baixo), tais como, memória de curto prazo, memória episódica, conhecimento previamente adquirido sobre cenas visuais etc., tem sido considerada em um conjunto de trabalhos.

Segundo Henderson (2003), o controle do olhar pode ser influenciado pelos estímulos visuais e/ou pelo conhecimento do observador. Em tarefas em que um participante seja exposto a novos conhecimentos, a saliência visual pode dar lugar ao aprendizado de novos dados sobre relações entre os elementos visuais observados. Além disso, em tarefas complexas e aprendidas como ler e cozinhar, o olhar do observador assume características específicas (p. 501).

Maia (2008) investigou a interação entre processos *bottom-up* e *top-down* na cognição de elementos visuais com 27 indivíduos adultos, falantes de português brasileiro. O pesquisador utilizou cópias digitais de uma tela ("Idílio") de Tarsila do Amaral que foram manipuladas para retirar ou acrescentar certos elementos visuais. Buscou verificar se a apresentação de um título previamente à imagem poderia influenciar o processo de visualização da mesma. Os participantes viram imagens precedias pelos slides com títulos "Casa na colina", "Pedras no riacho" e imagem sem título. As telas com as imagens eram manipuladas da seguinte forma: uma das telas mostrava um casal e uma casa; outra, um cão e uma casa, e a terceira, uma casa. Todas as telas apresentavam pedras e desenhos representando árvores e plantas. As variáveis independentes foram, portanto, propriedades do estímulo (casal, cão, nada) e o título prévio (casa, pedras, nada). As variáveis dependentes foram os tempos totais de fixação do olhar e os movimentos sacádicos iniciais (medidas *on-line*) e um relato, de um parágrafo, que os participantes produziram após visualizarem a imagem (medida *off-line*).

Como resultados, Maia (2008) observou que os títulos prévios (fator de *top-down*) influenciam o relato dos participantes, mas não se sobrepõem às propriedades estruturais salientes (*bottom-up*) no processo de rastreamento ocular da imagem. Por exemplo, um efeito *bottom-up* foi verificado quando os participantes liam o título "Pedras no riacho" e viam a tela com o casal. A imagem do casal mostrou-se saliente, de modo que atraiu mais fixações iniciais, levando-se à con-

clusão de que fatores *top-down* têm influência nos movimentos sacádicos, porém certas características, como traços de animacidade, podem gerar um efeito *bottom-up*. Os dados obtidos no experimento indicaram, ainda, haver uma hierarquia na busca visual determinada pelo traço de animacidade.

2.2
A caracterização da visão segundo Marr (1982)

Não se pode discutir o reconhecimento de objetos sem fazer menção a David Marr, que publicou a obra *Vision*, em 1982, e a Irving Biederman, que, em 1987, apresentou um aprofundamento dos estudos de Marr. Este autor deu grande contribuição às ciências cognitivas analisando a visão de maneira interdisciplinar – a anatomia e a fisiologia não seriam suficientes para explicar certas características dos neurônios, sendo necessário recorrer a princípios matemáticos, conforme salienta Gardner (2003, p. 314).

Marr (1982) abordou a visão como um sistema de processamento de informação. O autor define a visão como "um processo que, a partir de imagens do mundo externo, produz uma descrição útil ao observador e que não é repleta de informação irrelevante" (Marr, 1982, p. 31). O autor propôs três diferentes níveis de análise a partir dos quais se poderia estudar a visão: o nível computacional, o nível algorítmico/representacional e o nível da realização física. No nível computacional, o sistema realiza o mapeamento de informações, de modo que as propriedades abstratas de tal mapeamento sejam precisamente definidas. Nesse estágio, o objetivo da computação, sua adequação e lógica estratégica devem ser estabelecidas a fim de atender a uma finalidade. Em seguida, no nível representacional/algorítmico, define-se como a computação será implementada no sistema, estratégias para lidar com dados de entrada e saída do sistema e com base em que algoritmo eles serão transformados. Finalmente, no nível da implementação, busca-se caracterizar a realização física da representação e do algoritmo. O autor enfatiza que existe uma relação causal e lógica entre esses níveis, ainda que não muito fixa. Certos fenômenos poderiam ser explicados por apenas dois desses níveis, de modo que certas considerações psicofísicas sejam feitas apropriadamente.

A fim de explicar o processamento de representações simbólicas como um mapeamento de uma representação para outra, o autor considera a imagem como uma representação, que deve ser, essencialmente, útil para o observador. Como a visão é utilizada entre os animais com uma grande variedade de propósitos, o autor não acredita que as representações utilizadas pelos animais sejam as mesmas. Logo, elas devem variar conforme o propósito de cada um deles. Ainda que a visão humana seja muito mais complexa que a de diversos animais, estudos mencionados pelo autor, como o da visão de um tipo de mosca, que foi explicitado por equações matemáticas, podem explicitar quais conexões neurais seriam responsáveis pela visão humana.

O autor considera impossível que o resultado do processamento visual ocorra em apenas um estágio. Para ele, o sistema de processamento visual lida com uma série de representações, inicialmente mais grosseiras e, posteriormente, mais precisas das propriedades físicas de um objeto. O esboço primordial, o esboço em 2 ½-D e o modelo de representação 3-D corresponderiam aos resultados dos estágios de processamento. Os esboços corresponderiam a representações iniciais do processamento visual, que guardariam dados básicos dos objetos observados, e que contribuiriam para um modelo representacional em 3-D, mais robusto e completo. Gardner (2003) explica que

> ao estabelecer estes esboços, Marr e seus colegas estavam traçando os passos através dos quais qualquer mecanismo necessariamente passa do momento (ou circunstância) em que ele tenta pela primeira vez tornar uma cena externa inteligível, até o momento (ou circunstância) em que a cena foi apreendida de forma relativamente verídica (p. 317).

O esboço primordial enfatiza dados obtidos a partir de extremidades, contornos e saliências dos objetos. O esboço em 2 ½-D considera dados de sombreamento, textura, profundidade, movimento etc., conforme o ponto de vista do observador. A representação do modelo de 3-D descreve os objetos "independente do ponto de vista do observador (portanto, invariável em termos do ponto de vista)" (Marr, 1982 *apud* Eysenck & Keane, 2007, p. 79). Biederman (1987) amplia a teoria de reconhecimento de objetos de Marr, propondo que os objetos seriam formados por componentes menores, os *geons* (ícones geométricos). Segundo Biederman, o observador poderia identificar qualquer objeto com base em quais-

quer representações deste se forem adequadas às informações do componente ou do *geon* do objeto visual (Eysenck & Keane, 2007, p. 83). Uma questão fundamental no reconhecimento de objetos é o problema da ligação (*binding problem*): como se dá a integração de diferentes fontes de informação a fim de reconhecer objetos. Hummel & Biederman (1992 *apud* Eysenck & Keane, 2007, p. 84), de modo a responder essa questão, sugeriram uma abordagem conexionista da teoria dos *geons*. De acordo com esse modelo, as unidades da rede seriam ativadas em grupos vinculados. Entretanto, o modelo é limitado, pois foi testado com um número restrito de objetos, ainda que tais objetos tenham sido testados em orientações diferentes. Esse processo de reconhecimento de objetos foi feito com detalhes suficientes para que um computador pudesse simular os passos que o cérebro humano segue para computar um percepto[1] em 3-D daquilo que é visto.

As teorias de Marr e de Biederman propõem uma hierarquia nos processos de reconhecimento de objetos. Riddoch & Humphreys (2001, *apud* Eysenck & Keane, 2007, p. 91), ao analisar pacientes com agnosia visual, um tipo de deficiência que afeta o reconhecimento de objetos, propuseram o seguinte modelo hierárquico:

> - Agrupamento das bordas por co-linearidade: é um estágio inicial do processamento, durante o qual as bordas de um objeto são derivadas (co-linear significa ter uma linha comum).
> - Ligação de traços e formas: durante este estágio, os traços que foram extraídos do objeto são combinados para compor formas.
> - Normalização do ponto de vista: durante este estágio, o processamento ocorre para permitir que seja derivada uma representação independente do ponto de vista. Este estágio é controvertido, pois muitas evidências sugerem que o reconhecimento de objetos nem sempre envolve representações independentes do ponto de vista.
> - Descrição estrutural: durante este estágio, os indivíduos obtêm acesso ao conhecimento armazenado sobre as descrições estruturais do objeto.

[1] De acordo com o "Oxford Dictionary" (www.oxforddictionaries.com), um *percepto* é um objeto da percepção, algo que é percebido, um conceito mental resultante do processo de percepção. O dicionário "Merriam-Webster" (www.merriam-webster.com) relaciona *percepto* à expressão *sense-datum*, um objeto particular imediato da sensação que não é passível de análise, distinto do objeto em si. A definição de *sense-data*, segundo a "Stanford Encyclopedia of Psychology", apresenta correntes do pensamento filosófico que defendem haver semelhanças entre as propriedades do objeto e de seu percepto e correntes que são contrárias a essa ideia.

> • Sistema semântico: o estágio final no reconhecimento de objetos envolve obter acesso ao conhecimento armazenado das informações semânticas relevantes para um objeto. (Eysenck & Keane, 2007, p. 91)

No que tange ao estágio de acesso a informações semânticas, Hills & Caramazza (1991) estudaram o caso de dois pacientes com déficits específicos de categoria, que afetam o reconhecimento de certos tipos de objetos – um deles tinha dificuldades em identificar seres animados e o outro, inanimados. Humphreys & Forde (2001) sugeriram que a dificuldade em identificar seres animados pode estar relacionada à maior semelhança existente entre seres animados. Conforme Caramazza (2001, p. 441 *apud* Eysenck & Keane, 2007, p. 95), é possível que existam diferentes áreas do cérebro dedicadas ao reconhecimento de seres animados e inanimados. Por exemplo, o processamento de objetos animados tem sido associado às áreas inferiores do lobo temporal e o processamento de objetos inanimados, às áreas frontoparietais. As evidências comprobatórias da especificidade de áreas do cérebro dedicadas ao reconhecimento de seres animados e inanimados foram obtidas a partir de estudos com pacientes com lesões cerebrais nas áreas previamente mencionadas.

2.3
Percepção e representação de estímulos visuais

A percepção e a representação de estímulos visuais têm sido discutidas sob diferentes perspectivas, assim como a descrição do processamento de informações de ordem imagética e linguística. Nesta seção, as teorias da dupla codificação (*Dual Coding Theory*) de Allan Paivio (1991) e a teoria proposicional, especialmente discutida por Pylyshyn (1978, 1981, 2003) são apresentadas.

2.3.1
A teoria da dupla codificação

Allan Paivio desenvolveu a teoria da dupla codificação (*Dual Coding Theory*, ou "DCT" na sigla em inglês) a partir de experimentos de aprendizado associativo (Paivio, 1991). Paivio acredita que a DCT conseguiu muita atenção rapidamente por ter sido a primeira investigação sobre o estudo da imagem e suas funções (Paivio, 1991, p. 256).

Os testes iniciais envolvendo medições da imagem em relação à memória procuravam lidar com "(a) o valor da evocação da memória em estruturas linguísticas maiores; (b) procedimentos experimentais elaborados para encorajar ou interferir com o uso da imagem; e (c) diferenças individuais em habilidades com imagem"[2] (1991, p. 256). O autor concluiu que as informações verbais e imagéticas são processadas de forma diferente por meio de testes de memorização em que os participantes eram expostos a sequências de palavras e imagens e depois tinham de recordar o que viram ou leram em diferentes ordens. Os participantes, no relato das realizado, demonstraram mais facilidade em recordar os itens em ordem aleatória, tendo sido as sequências de palavras mais bem relembradas que as sequências de imagens (Sternberg, 2008, p. 228). Paivio sintetiza o debate sobre a DCT da seguinte forma:

> A DCT se baseia na assunção de que o substrato representacional é multimodal e relativamente concreto, ao passo que a posição oposta dominante tem sido a de que a linguagem do pensamento é unimodal e abstrata, classicamente vista como palavras internalizadas e sentenças, e, mais recentemente, como estruturas proposicionais amodais e processos governados por regras (computacionais). (Paivio, 1991, p. 256-257, tradução nossa)[3]

A fim de apresentar um panorama da DCT, Paivio faz uma distinção entre sistemas simbólico e sensório-motor. O autor acredita em uma relação ortogonal entre sistemas simbólicos e sistemas sensório-motores específicos. Para Paivio (1991, p. 257), os sistemas verbal e não-verbal representam simbolicamente as propriedades estruturais e funcionais da linguagem e do mundo não-linguístico, respectivamente. Ambas as "classes de eventos", como define o autor, apresentam-se em diferentes "modalidades":

- Visual – palavras impressas *versus* objetos visuais;
- Auditiva – palavras faladas *versus* sons do ambiente;
- Háptica – respostas táteis e motoras da escrita e como se relacionam à manipulação de objetos.

[2] O texto original não deixa claro como seriam os testes de medições da imagem em relação à memória e não apresenta exemplos de testes.

[3] Whereas DCT rests on the assumption that the representational substrate is multimodal and relatively concrete, the dominant opposing position has been that the language of thought is unimodal and abstract, classically viewed as internalized words and sentences and, more recently, as amodal propositional structures and rule-governed (computational) processes.

Paivio argumenta que a relação ortogonal é incompleta porque as modalidades gustativa, olfativa e afetiva não são verbais. Seria impossível construir símbolos a partir de sabores, odores ou emoções.

Assume-se que as unidades representacionais dos sistemas simbólico e sensorial possuam análogos perceptual-motores específicos para cada modalidade. Originalmente, Paivio referia-se a eles como representações verbais e imagéticas (no original em inglês, *imaginal*). O autor elegeu os termos *logogen* (gerador de palavras) e *imagen* (gerador de imagens) a fim de distinguir as representações geradas por eles. Sternberg (2008, p. 226-229) interpreta a teoria de Paivio associando as imagens mentais a códigos analógicos, que reteriam características perceptuais essenciais do estímulo visual. As representações mentais de palavras, por outro lado, seriam agrupadas em um código simbólico, uma representação abstrata, selecionado arbitrariamente, sem a obrigação de manter semelhança com o estímulo do ponto de vista perceptual. Os sistemas verbal e imagético funcionam de modo independente e podem, hipoteticamente, complementar-se.

A seguir apresenta-se um esquema das representações verbais e imagéticas propostas no âmbito da teoria da "dupla codificação" de Paivio (com base em Clark & Paivio, 1991):

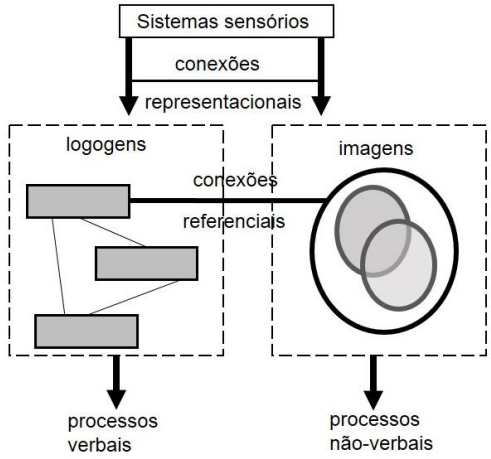

Figura 2 – Esquema de representação de estímulos verbais e visuais segundo a teoria da dupla codificação (adaptado de Clark & Paivio, 1991, p. 152)

Paivio (1991) enuncia algumas propriedades gerais dos sistemas simbólicos, como a independência funcional dos subsistemas, as conexões entre unidades, as operações de processamento e os processos organizacional e transformacional. A independência funcional dos subsistemas caracteriza-se pelo modo independente

como os sistemas verbal e não-verbal podem funcionar, ao mesmo tempo (em paralelo) ou em separado (um funciona e outro, não). As conexões entre unidades seriam necessárias para o funcionamento geral dos sistemas. Paivio explica que *imagens* e *logogens* estariam interconectados de modo que se possa nomear objetos e que, em direção oposta, nomes evoquem imagens. As interconexões estariam dispostas de um ponto para outros, em ambas as direções, supondo uma correspondência entre nomes e objetos. Talvez um objeto evoque muitos nomes, assim como talvez um nome evoque muitos referentes. A ativação é definida probabilisticamente, conforme a força das interconexões e o contexto, por meio das operações de processamento. Os processos organizacional e transformacional estão baseados na teoria de que cada sistema restringe a organização e a transformação da informação a seu modo. Segundo o autor:

> O sistema verbal gera estruturas sequenciais em diferentes níveis (sintagmas, sentenças) restritas pelo "*sequential frame*", de modo que a ordem temporal dos elementos possa ser mudada ou novos elementos substituídos por outros que ocupem uma posição temporal. Transformações não-verbais são governadas por restrições estruturais e de processamento associadas a representações não-verbais. Então, elas incluem transformações em qualquer dimensão espacial (e.g., rotações mentais, mudanças no tamanho ou forma das imagens etc.), mudanças em propriedades sensoriais (cor imaginada, qualidade sonora etc.), movimento dos objetos imagéticos e assim por diante. (*Idem*, p. 260, tradução nossa)[4]

A hipótese *conceptual-peg*, relacionada ao efeito das imagens e palavras no momento da recordação, é o fundamento da DCT (p. 260). Segundo o autor, essa hipótese origina-se de listas de itens para testes mnemônicos (*memory-peg tests*). A codificação dupla é necessária à geração de palavras durante o aprendizado de listas. A técnica implica processos de DCT do tipo (a) processamento de referência imagética/verbal; (b) processamento verbal associativo (baseado parcialmente em relações de rima); (c) organização das imagens (integração). Os componentes verbal e imagético corroboram a retenção de informação na memória.

[4] The verbal system generates sequential structures of different levels of complexity (e.g., phrases, sentences) and transformations are constrained by that sequential frame, so that the temporal order of elements can be changed or new elements substituted for ones that occupy a particular temporal slot. Nonverbal transformations are governed by the structural and processing constraints associated with nonverbal representations. Thus, they include transformations on any spatial dimension (e.g., mental rotations, changes in the size or shape of images, etc.), movement of imaginal objects, and so on.

2.3.2
A teoria proposicional

Pylyshyn (1978, 1981, 2003) e Anderson & Bower (1973, *apud* Sternberg, 2008) foram os maiores expoentes da teoria proposicional, que se apresenta como uma alternativa à DCT. Segundo Sternberg, a teoria proposicional postula que as representações mentais não seriam armazenadas sob a forma de imagens, porém de modo mais abstrato. Enquanto uma proposição expressa relações entre conceitos – sentidos subjacentes a relações, uma imagem é um epifenômeno, isto é, resultante de processos cognitivos secundários.

Sobre a noção de proposição, é importante ressaltar que não há um acordo entre pesquisadores de áreas como a inteligência artificial, a lógica, a linguística e a psicologia cognitiva: cada uma dessas áreas assume uma dada definição de proposição. Field (2004), em um glossário na área de Psicolinguística, define proposição como "uma representação abstrata de uma única unidade de sentido: um registro mental do significado central de uma sentença sem quaisquer dos fatores interpretativos e associativos que o leitor/ouvinte possa trazer a ele" [5]. (p. 225)

Proposições podem ser expressas a partir de um cálculo predicativo, um artifício criado para indicar sentidos subjacentes a relações, no qual as diferenças superficiais são desconsideradas para favorecer as mais profundas. A expressão lógica do cálculo predicativo apresenta-se da seguinte forma:

[Relação entre elementos] ([Elemento sujeito], [Elemento objeto])

Por exemplo, Sternberg (2008, p. 230) cita uma relação que envolvesse ação do tipo "um camundongo mordeu um gato", que poderia ser representada sob a forma "Morder [ação] (camundongo [agente da ação], gato [objeto])". Relações espaciais do tipo "um gato está debaixo da mesa" seriam representadas como "[posição verticalmente mais elevada] (mesa, gato)". Dessa forma, as informações de ordens visual (imagens) e linguística (verbais) seriam armazenadas sob a forma de proposições. No momento em que se necessita recuperar na memória o dado armazenado, a mente poderia recuperar o código verbal ou de imagem (p. 231).

[5] An abstract representation of a single unit of meaning: a mental record of the core meaning of the sentence without any of the interpretative and associative factors which the reader/listener might bring to bear upon it.

Pylyshyn (1978) caracteriza as representações mentais partindo das seguintes etapas: "o que acontece quando uma cena é percebida e é assimilada em nosso conhecimento e o que acontece quando acessamos esse conhecimento mais tarde via memória ao relembrar tal cena" (p. 20). Fazendo referências a Julesz (1971), que discutiu a "visão ciclopeana", a Hochberg (1968), que comentou sobre o "olho da mente" e a David Marr (1975), que investigou as computações necessárias a esse tipo de visão, o autor caracteriza um conversor ("*transducer*", em inglês), da seguinte maneira:

> (a) Existe uma fase semiautônoma e pré-atencional na percepção visual; (b) essa fase é iniciada por energia chegando aos órgãos sensoriais; (c) somente o resultado dessa fase, e não as etapas intermediárias, estão disponíveis para análise perceptual mais aprofundada e (d) tais processos cognitivos como a percepção e assimilação de padrões sensoriais na forma de estruturas cognitivas ocorrem após essa fase. (Pylyshyn, 1978, p. 21, tradução nossa)[6]

O autor defende um ponto de vista computacional para esse momento semiautônomo da visão, em que a visão funcionaria como um conversor. A partir desse estágio, ocorreria a transformação do produto da conversão em representação na memória. O autor lista pontos importantes sobre esse construto mental:

1. A transformação do produto da conversão em memória é algo mais que uma simples degradação. Não se percebe ou se lembra de algo completamente em todos os aspectos, porém em baixo nível de detalhe ou precisão.
2. Motivação, expectativas, conhecimento prévio e estágio do desenvolvimento cognitivo influenciam a percepção e, consequentemente, a representação mental gerada a partir dela.
3. Existem muitos aspectos que não se enquadram como sensoriais ou pictóricos, como relações de causa e efeito, ataque e defesa e relações espaciais. Tais relações conceptuais são abstratas e não seriam analisadas pelo conversor.
4. A sequência de etapas referente ao processo de relembrar uma imagem ou cena e perceber seus aspectos é problemática. A recuperação de uma

[6] (a) There is a semi-autonomous, pre-attentive phase in visual perception, (b) this phase is initiated by energy arriving at the sense organs, (c) only the output of this phase, and not intermediate steps, are available for further perceptual analysis, and (d) such cognitive processes as "noticing" and the assimilation of sensory patterns into cognitive structures take place after this phase.

> cena ou parte dela pela memória acessando aspectos do conteúdo perceptualmente interpretado baseia-se em dados armazenados que foram previamente interpretados sem necessidade de re-percepção. A retenção de imagens obedece a uma hierarquia de detalhes e aspectos ilimitados. (p. 22-23).

Pylyshyn (1978) acredita que o termo "imagem" pode levar a interpretações imprecisas. Para ele, não se deve pensar a representação como uma imagem. O autor prefere utilizar os termos "descrição" ou "descrição estrutural", que trariam conotações desejáveis. O termo "descrição" pode ser associado a algo construído com base em conceitos encontrados, por exemplo, nas "categorias do conhecimento" de Kant, pode trazer uma relação referencial com o objeto representado em vez de uma relação de "semelhança" e tem semântica "definida por uma função de acesso que não se pode considerar como o aparato visual completo" (p. 24).

Pylyshyn também faz uma crítica à ideia de que as representações mentais são analógicas. A correspondência entre as operações "no mundo" e as operações mentais é parcial – apenas alguns aspectos de algumas operações físicas têm correspondências.

Pylyshyn (2003), argumentando a favor de uma representação não pictorial, conclui que, "uma vez que a informação entre no sistema visual (de maneira oposta a permanecer na retina), ela não parece funcionar como os inputs visuais, no que concerne a mostrar propriedades 3-D" (p. 1-17)[7]. Uma cena contém muito mais detalhes geométricos que a informação dela extraída. Os conceitos visuais são abstratos e variáveis em detalhes. Fazendo um paralelo com a linguagem, somos capazes de descrever uma cena com muitos detalhes ainda que não mencionemos a localização de alguns objetos ou ofereçamos dados superficiais de sua forma.

As figuras abaixo ilustram a ideia de como poderiam ser as representações obtidas a partir da visualização de uma cena:

[7] (…) Once the information gets into the visual system (as opposed to still being on the retina) it no longer seems to function the way visual inputs do, in terms of showing such signature properties as automatic three-dimensional interpretation and spontaneous reversals.

> (...) A informação sobre uma cena visual não é armazenada em uma forma pictorial, mas em uma forma de descrição, que é caracterizada por padrões e abstrações variáveis e é baseada em conceitos disponíveis. Então, em vez de pensar na visão como no *cartoon* de Kliban na primeira figura, deve-se substituir a figura no balão de pensamento por uma estrutura de dados como na segunda figura, em um formato que é tipicamente usado em aplicativos de inteligência artificial. (Pylyshyn, 2003, p. 1-21)[8]

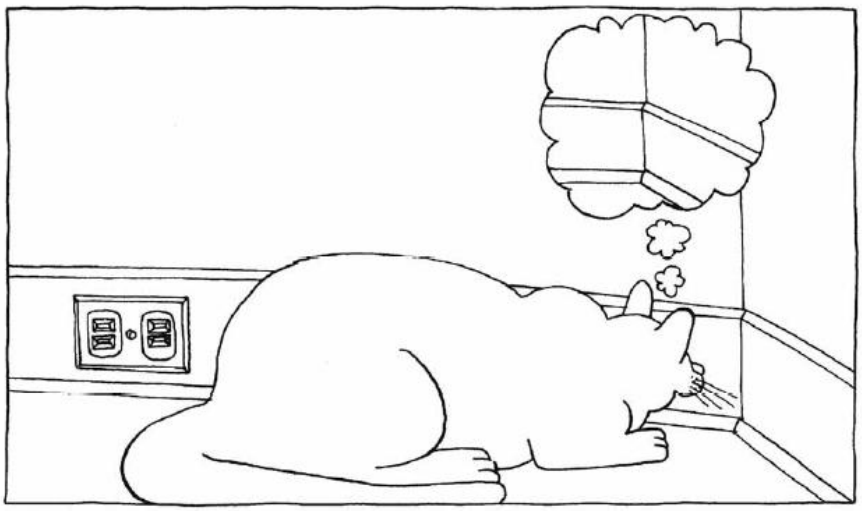

Figura 3 – Concepção de representação mental imagética criticada por Pylyshyn (2003).

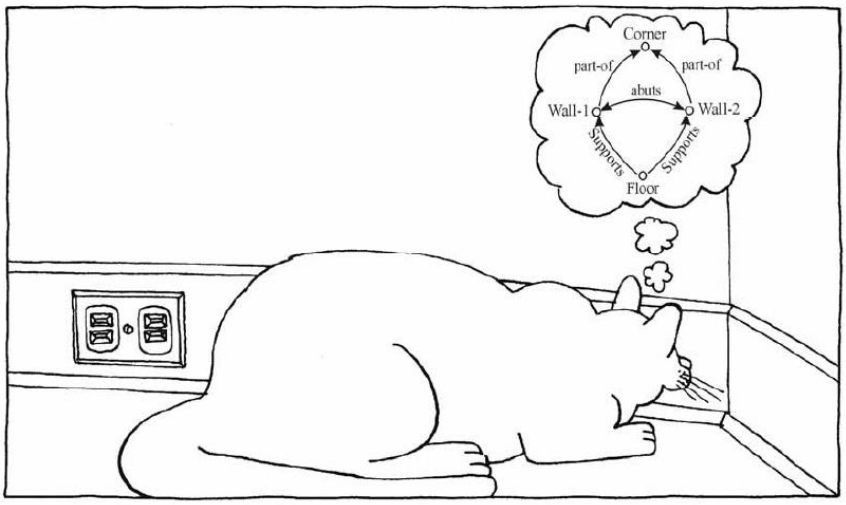

Figura 4 – Sugestão de representação mental (conteúdo proposicional) que seria armazenada na memória de acordo com Pylyshyn (2003).

O conteúdo de imagens e perceptos visuais poderiam ser, portanto, codificados na forma de proposições. Isso explicaria melhor as propriedades conceptu-

[8] (…) information about a visual scene is not stored in a pictorial form, but rather is stored in a form more like that of a description, which is characterized by variable grain and abstractness and is based upon available concepts. Thus rather than thinking of vision the way it was depicted in the Kliban cartoon in Figure 1-1, one should replace the picture in the thought balloon (shown on the left panel of Figure 1-20) with a data structure such as on the right panel, in a format that is typically used in artificial intelligence applications.

ais de imagens mentais e da própria percepção (cap. 7, p.30-31). Contudo, o autor ressalta ser necessária uma teoria que explicite em mais detalhes como fenômenos imagéticos seriam codificados proposicionalmente. "A ideia de que imagens são codificadas na forma de proposições é uma *hipótese nula* (grifo do autor) contra a qual outras propostas sobre as representações de imagens devem ser testadas" (*Idem,* p. 7-31). O conhecimento recentemente obtido sobre proposições a partir da lógica formal, e da teoria da computação pode contribuir para desenvolver teorias da codificação de imagens mentais.

2.3.3
Atenção visual e linguagem

Nesta seção, algumas questões referentes a processos atencionais sob uma perspectiva da psicologia cognitiva serão discutidas, enfatizando como se pode relacioná-los a tarefas envolvendo estímulos linguísticos e visuais, uma vez que os experimentos realizados neste trabalho exigiram a atenção visual dos participantes. A atenção vem sendo há tempos tema de discussão da psicologia cognitiva e se relaciona com o grau de seletividade do pensamento (Eysenck & Keane, 2007, p. 142).

Eysenck & Keane (2007) relatam restrições importantes na pesquisa sobre atenção. Apesar de um indivíduo ser capaz de fixar sua atenção no ambiente interno (entendido como os pensamentos próprios e fatores como memória de longo prazo) ou externo, a maior parte da pesquisa concentra-se no ambiente externo, pois é mais simples controlar estímulos nele. Durante a execução de uma tarefa, um indivíduo tende a focalizar sua atenção em seu objetivo; isso, no entanto, não garante que fatores internos, como motivação, não provoquem interferência nos resultados. As pesquisas realizadas até o momento indicam, contudo, que aspectos motivacionais têm menos relevância que instruções dos experimentos.

Silva *et al.* (2011, p. 27) fazem referência a duas formas de direcionamento atencional – a orientação voluntária, que se dá em função de um controle *top-down*, e a orientação automática, que se dá em função de estímulos inesperados, oriundos do ambiente, e cujo processamento é do tipo *bottom-up*. Os dois modos de direcionamento coexistiriam de modo que a atenção é um produto da competi-

ção desses dois direcionamentos (Berger, Henik, Rafal, 2005 *apud* Silva *et al.*, 2011).

A caracterização do que é selecionado em um processo atencional tem sido feita segundo três perspectivas: a atenção seletiva a uma área específica do espaço, a atenção a determinado(s) objeto(s) e a alternância entre a atenção a uma área do espaço e a um objeto (Eysenck & Keane, 2007, p. 142). De maneira similar a um holofote, todos os elementos presentes em uma pequena área do campo visual podem ser vistos. Posner (1980, p. 5), supondo a ideia do holofote e considerando movimentos da cabeça e dos olhos, propôs a noção de atenção encoberta (*covert orienting*), segundo a qual o holofote da atenção pode localizar-se em posição diferente caso não haja movimento dos olhos. Em um dos experimentos realizados, os participantes deveriam pressionar um botão independentemente da posição do estímulo (um quadrado preto) no display. Os participantes viram um sinal de mais ou uma seta apontando para a esquerda ou para a direita (pistas de atenção visual) no centro da tela. Se o sinal de mais aparecesse, a detecção do estímulo poderia ocorrer à esquerda ou à direita da fixação ocular (somente os tempos de resposta dos participantes que, no início da tarefa, fixaram o olhar na pista de atenção no centro do display foram considerados). Se uma seta aparecesse, o estímulo poderia aparecer no lado indicado pela seta (detecção de estímulo válida) ou no lado oposto (detecção de estímulo inválida) (Posner, 1980, p. 6-7).

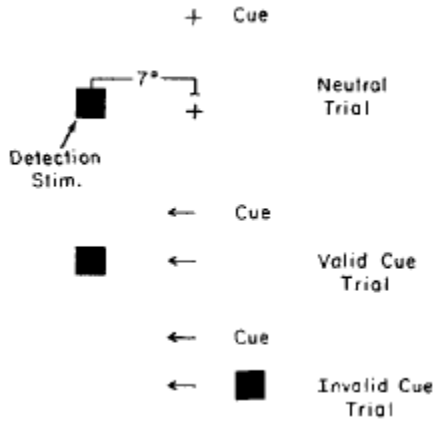

Figura 5 – experimento com pista de atenção central de Posner (1980, p.6).

Como resultados, percebeu-se que, nas situações em que a seta indicava a posição correta do estímulo (detecção válida), a resposta do participante era mais rápida do que quando não havia indicação ou quando a dica era incorreta. Esse

estudo corrobora a noção de que pistas de atenção interferem no comportamento de participantes. Myachykov *et al.* (2009, p. 3) observam o trabalho de Posner (1980) como relevante contribuição para o paradigma das pistas visuais (*cueing paradigm*).

Tomlin (1997, p. 173) descreve a atenção tendo por base um modelo visual, que compreenderia três processos independentes, o estado de prontidão (*alerting*), a orientação e a detecção. O estado de prontidão relaciona-se a novos estímulos; a orientação compreende como o sistema consegue selecionar certos estímulos em detrimento de outros; a detecção relaciona-se como se dá a seleção e o registro de um item para posterior processamento.

Tomlin (1995) investigou voz verbal em inglês fazendo uso de estímulos visuais. Os participantes, falantes nativos de inglês, foram expostos a uma sequência de animação de eventos dinâmicos, conhecida como "The fish film" (ver figura abaixo), que tinham de ser descritos de modo *on-line*, isto é, assim que eram visualizados. Em cada animação viam-se dois peixes nadando um em direção ao outro; em seguida, um peixe engolia o outro. A atenção dos participantes foi manipulada por meio de uma seta que aparecia logo acima de um dos peixes. As cores dos peixes foram selecionadas aleatoriamente. Em metade dos estímulos, o elemento agente era salientado pela manipulação de atenção, e na outra metade, o elemento paciente era salientado. A manipulação de atenção era do tipo explícita, uma vez que os participantes foram instruídos a focalizarem sua atenção no peixe que fosse marcado pela seta e, terminada a animação, a descreverem a cena. Sentenças do tipo "The red fish ate the blue fish" e "The blue fish was eaten by the red fish" foram eliciadas. Esperava-se que o elemento manipulado assumisse o papel de sujeito na produção dos participantes. Nas tarefas em que o elemento agente era destacado, esperavam-se sentenças na voz ativa, e nas tarefas em que o elemento paciente era destacado, esperavam-se sentenças na voz passiva. Como resultado, a produção dos participantes aproximou-se das previsões. De um total de 12 participantes, apenas dois descreveram as cenas utilizando somente voz ativa, o que pode estar relacionado a fatores como a saliência perceptual na animação no momento em que um dos peixes abre a boca, ainda que a animação não tenha sido marcada pela manipulação de atenção, a maior frequência de estruturas na voz ativa e inclusive a fatores endógenos dos participantes. Tomlin (1995, p.

529) relata que um desses dois participantes, em entrevista após ter realizado a tarefa, disse que evitou produzir sentenças na voz passiva por vontade própria. O autor concluiu que falantes de inglês tendem a relacionar um elemento focalizado primeiro no momento da formação da sentença ao papel de sujeito em sua produção linguística. Esse material visual foi utilizado em outro experimento com falantes de ordens mais flexíveis que o inglês e os resultados foram diversos (Cf. 4.3.3).

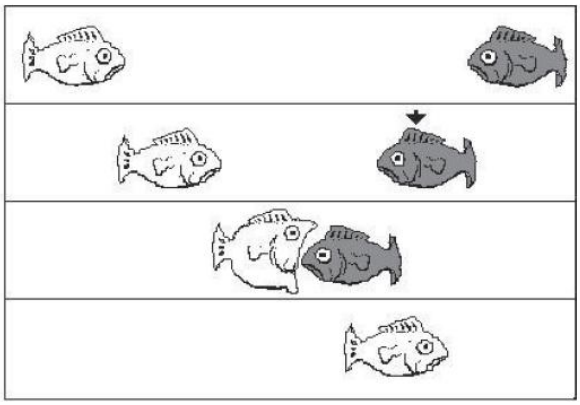

Figura 6 – Sequência de animação “The Fish Film” com manipulação de atenção (seta) utilizada por Tomlin (1997).

Considerando o experimento de Tomlin (1997), Gleitman *et al.* (2007) (Cf. 4.3.2) desenvolveram experimentos com manipulação de atenção visual que influenciaram os experimentos realizados para esta pesquisa. No entanto, conforme será visto no capítulo 4, as pistas de manipulação de atenção utilizadas por Gleitman *et al.* (2007) foram implícitas, ou seja, bem menos evidentes que as de Tomlin (1997); as pistas de manipulação de atenção utilizadas nos experimentos realizados para esta pesquisa também foram do tipo implícita.

No próximo capítulo, serão mais diretamente abordadas questões relativas à interface linguagem-visão, a partir de uma breve apresentação sobre a ideia de modularidade da mente e da discussão de como dados de ordem visual e linguística poderiam ser mentalmente integrados. Será retomada a ideia de representações imagéticas e proposicionais, no contraste entre a hipótese de módulos mentais híbridos (Jackendoff, 1996, 2002) e uma proposta alternativa, alinhada à visão de Pylyshyn de que representações proposicionais seriam geradas tanto a partir de estímulos visuais como verbais.