4
Interação linguagem-visão: estudos psicolinguísticos

4.1
Apresentação

Ferreira & Tanenhaus (2007), em texto introdutório a um número especial do "Journal of Memory & Language" acerca das interações entre linguagem e visão, comentam que ainda são poucos os trabalhos que exploram como se dá a integração entre os sistemas linguístico e visual no processamento de informações. Isso se observa mesmo nos casos em que os estudos em um domínio fazem uso de técnicas experimentais em que as respostas dos participantes envolvem o outro domínio. No caso de estudos de percepção visual, muitas vezes respostas verbais são tomadas como variáveis dependentes em um dado experimento. No caso dos estudos sobre processamento linguístico, por sua vez, o emprego de gravuras é muito comum em um conjunto de técnicas experimentais empregadas tanto na produção quanto na compreensão (2007, p. 455).

Um aspecto bastante grave é que, tanto de um lado quanto de outro, pouco se sabe sobre o outro domínio. Na área da visão, aspectos, por exemplo, associados à organização do léxico mental e acesso lexical não são considerados quando se analisam respostas linguísticas (tempo de nomeação de um estímulo visual, por exemplo) como informativas sobre algum aspecto da percepção visual. Do lado dos psicolinguistas, por sua vez, considera-se que medidas relativas a fixações e movimentos sacádicos em tarefas em que frases são apresentadas concomitantemente a *displays* com objetos e imagens refletem de modo direto processamento linguístico. Note-se, contudo, que questões mais diretamente relacionadas à interface são pouco exploradas (p. 455-456).

Neste capítulo, serão apresentados alguns trabalhos em psicolinguística – tanto em produção quanto em compreensão, nos quais respostas relativas ao mapeamento visual foram tomadas como informativas sobre o processamento linguístico. Serão destacados, nesses trabalhos, pontos que podem ser relevantes para os objetivos da presente pesquisa. No que tange aos estudos de compreensão, serão focalizados os trabalhos que buscaram investigar processos preditivos a par-

tir do paradigma do mundo visual[1] e será comentado o trabalho de Knoerfele & Crocker (2006), em que se postula a teoria da Inter-Relação coordenada (*Coordinated Interplay Account*) entre informação linguística e visual. No caso da produção, serão comentadas pesquisas em que o movimento do olhar dos participantes em uma tarefa de descrição de cenas foi tomado como pista acerca da incrementalidade na passagem de informação do nível da mensagem para o nível da formulação sintática. Nesses últimos estudos, também será apresentado um artifício experimental que foi adotado em nossa pesquisa – o recurso de manipulação de atenção empregado por Gleitman *et al.* (2007). Na última seção, será referenciado o estudo realizado por Clark & Chase (1972), provavelmente um dos primeiros trabalhos em processamento no qual a questão da integração entre informação linguística e visual é diretamente enfocada, a partir da investigação de como se dá o processo de comparação entre sentença e figura.

4.2. Estudos relacionando compreensão de linguagem e informação visual

4.2.1. Processos preditivos e o paradigma do mundo visual

Nesta seção, serão reportados alguns estudos que buscaram investigar processos antecipatórios/preditivos no processamento de sentenças, em que se fez uso do paradigma do mundo visual como método experimental (Altmann & Kamide, 1999; Kamide *et al.*, 2003). Embora neste trabalho o foco não seja em processos preditivos, os estudos a serem reportados são relevantes por levantarem questões não apenas acerca do curso incremental do processamento, mas também porque proveem indicações sobre como se coordenam informação linguística e visual e como o processamento da linguagem pode afetar o processo de busca visual[2]. Conforme será explicado na metodologia dos experimentos, os estímulos linguístico e visual não foram apresentados ao mesmo tempo para os participantes, de modo diverso dos experimentos de Altmann & Kamide (1999) e Kamide *et al.*

[1] O paradigma do mundo visual foi iniciado por Cooper (1974) e, mais tarde, desenvolvido por Tanenhaus e colaboradores (1995). O paradigma do mundo visual consiste em experimentos em que um observador é exposto a informações linguística e visual e seus movimentos oculares são monitorados.

[2] O artigo "Anticipatory processes in sentence processing" (Kamide, 2008) apresenta uma ampla revisão de como processos antecipatórios têm sido investigados em psicolinguística, por exemplo, por meio de experimentos incluindo monitoramento ocular, a metodologia do paradigma do mundo visual e técnicas de neuropsicologia.

(2003), que investigaram processos preditivos/antecipatórios em tarefas em que os participantes eram expostos aos estímulos linguístico e visual concomitantemente.

Altmann & Kamide (1999) exploraram como a antecipação de informações poderia ocorrer com base em restrições de seleção semântica, em sentenças com verbos de dois argumentos. Por exemplo, o verbo *drink* (beber), de dois argumentos, exige um objeto direto (Tema) que deve ser algo bebível. Em seu primeiro experimento, na primeira condição, os participantes ouviram sentenças do tipo "*The boy will eat the cake*" (O menino vai comer o bolo), enquanto viam uma figura com um menino, um bolo, um trem de brinquedo, um carro de brinquedo e uma bola. O bolo era o único elemento na figura que poderia satisfazer as restrições de seleção do verbo "*eat*". Na segunda condição, os participantes ouviram sentenças do tipo "*The boy will move the cake*" (O menino vai mover o bolo) e viam a mesma figura (apresentada abaixo). Nessa condição, todos os elementos poderiam complementar o verbo "*move*". Altmann & Kamide (1999) puderam perceber que, na primeira condição, os movimentos sacádicos dos participantes eram muito mais direcionados para um dos elementos da figura que na segunda condição. Esses dados sugeriram que "o processador antecipa no verbo um argumento pós-verbal que o seguirá, aplica as restrições semânticas do verbo ao argumento e avalia o resultado considerando o contexto visual" (Kamide *et al.*, 2003, p. 135).

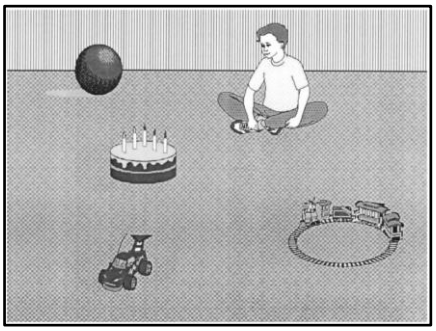

Figura 10 – Imagem exibida aos participantes quando ouviam "The boy will eat" ou "The boy will move" no experimento de Altmann & Kamide (1999).

Kamide *et al.* (2003) realizaram um estudo de compreensão de linguagem com monitoramento ocular aprofundando o trabalho de Altmann & Kamide (1999). No trabalho de 2003, os pesquisadores exploram processos preditivos relacionados à atribuição de papéis temáticos em contexto, por meio de dados relativos a movimentos oculares antecipatórios.

Os pesquisadores conduziram uma investigação com três experimentos – os dois primeiros, com falantes de inglês, e o terceiro, com falantes de japonês. No Experimento 1, trabalharam com verbos de três argumentos, contrastando sentenças do tipo "*The woman will spread the butter on the bread*" (A mulher vai passar manteiga no pão) e "*The woman will slide the butter to the man*" (A mulher vai passar a manteiga para o homem). Buscaram explorar como ocorreriam as antecipações de informações do segundo argumento pós-verbal (o Alvo) antes que ele fosse ouvido pelos participantes. No Experimento 2, investigaram como ocorreriam as antecipações em verbos de dois argumentos, porém considerando que o participante teria de fazer combinações entre os dados do Agente e do verbo, em sentenças do tipo "*The man will ride the motorbike*" (O homem vai andar de motocicleta) e "*The girl will ride the carousel*" (A menina vai andar no carrossel). No Experimento 3, os pesquisadores investigaram como dados morfossintáticos podem, por oposição a contingências do mundo real, influenciar antecipações durante o processamento de uma sentença. Sentenças do japonês com verbos de três argumentos foram analisadas, sendo que, em japonês, os argumentos do verbo sempre aparecem antes dele. O verbo ocorre em posição final, e cada argumento recebe marcação de caso (nominativo, acusativo, dativo).

No Experimento 1, os pesquisadores consideraram que, na língua inglesa, os verbos de três argumentos restringirão semanticamente seus alvos. Logo, eles previram que os participantes poderiam fixar os alvos corretos antes de ouvirem as preposições (no caso, "on" ou "to") nas sentenças "*The woman will spread the butter on the bread*" e "*The woman will slide the butter to the man*". Nas figuras, havia um agente, um tema, um alvo inanimado, um alvo animado e um elemento distrator. Como resultados, os pesquisadores puderam observar que, durante a apresentação do estímulo auditivo correspondente à expressão referencial pós-verbal (no caso, "the butter"), houve mais olhares direcionados aos alvos apropriados do que aos alvos inapropriados. Os movimentos oculares antecipatórios devem ter sido iniciados durante ou imediatamente após a referida expressão. Os autores não excluem, no entanto, uma análise alternativa desses resultados (como a proposta numa abordagem probabilística exemplificada por MacDonald *et al*., 1994), em que fatores associados à frequência de determinadas combinações argumentais poderiam explicar os efeitos antecipatórios obtidos.

No Experimento 2, os autores investigaram, em verbos de dois argumentos, como os participantes realizariam combinações dos dados semânticos, do agente e de restrições de seleção do verbo para antecipar seu complemento. O experimento foi concebido de modo a buscar investigar, de modo independente, contribuições (i) de informação combinatorial; (ii) informação sobre o agente; (iii) restrições selecionais do verbo. Para isso, os autores contrastaram sentenças do tipo "*The man will ride the motorbike*" (O homem vai andar de motocicleta) e "*The girl will ride the carousel*" (A menina vai andar no carrossel). No sentido de excluir possíveis efeitos de associações de baixo nível (*low level*) entre o agente e o tema na função de objeto, isto é associações independentes de restrições selecionais do verbo, os autores adicionaram duas condições experimentais exemplificadas pelas seguintes frases: "*The man will taste the beer*" (O homem vai saborear a cerveja) e "*The girl will taste the sweets*" (A menina vai saborear os doces).

Os resultados do experimento indicaram que o verbo combinado com o argumento pré-verbal pode restringir a antecipação de um tema subsequente. Houve um maior número de fixações no objeto correspondente à motocicleta na condição "*The man will ride*" do que na condição "*The girl will ride*", o que revela que o efeito antecipatório não foi determinado apenas por propriedades selecionais do verbo, mas sim pela combinação deste com o sujeito. O contraste entre as condições "*The man will ride*" vs. "*The man will taste*", com maiores fixações para "motorbike" no primeiro caso, permitiu mostrar que o efeito antecipatório não foi decorrente de uma mera associação entre "*man"* e *"motorbike"*, reforçando, pois, a contribuição de fatores combinatoriais.

No Experimento 3, os pesquisadores analisaram sentenças do japonês com verbos de três argumentos. O japonês é uma língua em que o verbo aparece em posição final, logo, todos os argumentos o precedem. Em "*syoojo-ga neko-ni sakana-o yatta*" (A menina (nominativo) gato (dativo) peixe (acusativo) deu, ou seja "A menina deu o peixe ao gato), cada argumento recebe marcação de caso. Assim, a informação extraída de menina (nominativo) e de gato (dativo) permite que o ouvinte entenda que a menina tem função de agente e o gato, de alvo; consequentemente, o ouvinte poderá antecipar, com esses dados, qual será o tema verbal e até mesmo qual será o verbo, levando em conta conhecimentos de mundo. Os participantes ouviam as sentenças e viam alternativas de agente, tema e alvo. Duas

condições foram elaboradas: dativa e acusativa, ou seja, os participantes ouviram sentenças com verbos bitransitivos (por exemplo, "*weitoresu-ga kyaku-ni tanosigeni habaagaa-o hakobu*", isto é, garçonete (nominativo) cliente (dativo) alegremente hambúrguer (acusativo) trazer – A garçonete vai trazer alegremente o hambúrguer ao cliente) e monotransitivos (por exemplo, "*weitoresu-ga kyaku-o tanosigeni karakau*", isto é, garçonete (nominativo) cliente (acusativo) alegremente perturbar – A garçonete vai alegremente perturbar o cliente). O padrão de fixações sugeriu que os participantes puderam prever que elemento poderia ser o tema mais plausível antes de ouvirem o início do termo correspondente. Dessa forma, os pesquisadores concluíram que a previsão de argumentos é possível mesmo em sentenças de uma língua de verbo na posição final. A marcação de caso auxiliou as antecipações, bem como uma análise dos argumentos pré-verbais.

Os três experimentos realizados por Kamide *et al.* (2003) propuseram uma reflexão sobre a incrementalidade no processamento de sentenças. "Para interpretar uma sentença de maneira incremental, palavra por palavra, é necessária a interpretação parcial do que foi encontrado até então – logo, depois de "*The boy will eat*", alguma representação deve ser construída (...)" (p. 152). Processos preditivos, ao anteciparem informação a ser processada, poderiam vir a contribuir para o processamento incremental de material linguístico.[3]

Para fins desta pesquisa, é relevante considerar que, em situações naturais de interação, material visual poderia funcionar como deflagrador de processos antecipatórios e que informação proveniente do processamento visual precisaria ser integrada à informação linguística. Uma forma de se pensar essa articulação será vista a seguir, com a apresentação da teoria da Inter-Relação Coordenada (*Coordinated Interplay Account* – CIA) da compreensão situada de sentenças, de Knoerfele & Crocker (2006).

[3] Como exemplo de trabalho em português envolvendo processos preditivos, a pesquisa de Forster (2013) pode ser mencionada. O autor conduziu um conjunto de experimentos com monitoramento ocular com a finalidade de detectar em que medida informações discursivas e contextuais poderiam ser acessadas durante o processamento on-line de sentenças com orações relativas restritivas de objeto. Nesse trabalho, a antecipação de referentes é explicada à luz de um modelo de computação on-line integrando elementos da Teoria Gerativa minimalista (cf. Chomsky 1995, 1999), e se assume um processador sintático autônomo. A pesquisa se insere em uma das linhas de investigação em decurso no LAPAL que tem por objetivo a conciliação entre processador e gramática, a investigação de questões de custos de processamento e de déficits de linguagem.

4.2.2
Teoria da Inter-Relação Coordenada (*Coordinated Interplay Account* – CIA) da compreensão

Knoerfele & Crocker (2006) realizaram estudos buscando relacionar a interação *on-line* entre mecanismos de compreensão de sentenças, conhecimentos linguístico e de mundo e processamento de cenas. Considerando a literatura, que vem apresentando resultados sugerindo uma relação temporal entre a atenção visual e a compreensão em tempo real, e sua própria pesquisa precedente, os pesquisadores determinaram duas dimensões fundamentais dessa interação: a dimensão temporal (*temporal dimension*) e a dimensão informacional (*informational dimension*). A dimensão temporal está relacionada à coordenação temporal entre a cena, a sentença e os conhecimentos linguístico e de mundo; a dimensão informacional "é a rápida influência de diversas fontes informacionais na compreensão incremental de sentenças"[4] (2006, p. 482).

Os autores realizaram experimentos utilizando técnica de monitoramento ocular com sentenças do alemão, que permite certa flexibilidade quanto à ordem de palavras, por exemplo, OVS, e cuja marcação de caso auxilia a eliminar a ambiguidade, de modo diverso do inglês, que tem a ordem SVO (sujeito-verbo-objeto) menos flexível. Nos dois experimentos, os participantes ouviram sentenças e visualizaram, ao mesmo tempo, figuras em que se viam três elementos animados correspondentes ao input linguístico.

O primeiro experimento, com sentenças do inglês, apresentou, nos resultados de monitoramento ocular, que, logo após terem ouvido o verbo, os participantes eram capazes de desfazer ambiguidades, que já seriam desfeitas antes mesmo de terem sido expostos à preposição ou ao segundo argumento (p. 498). Sentenças ambíguas do tipo "*The ballerina sketched by the fencer splashed the cellist*" (A bailarina desenhada pelo esgrimista molhou o violoncelista) e não ambíguas do tipo "*The ballerina splashed the cellist*" (A bailarina molhou o violoncelista) foram empregadas.

[4] (The second dimension) is the rapid influence of diverse informational sources on incremental utterance comprehension (…).

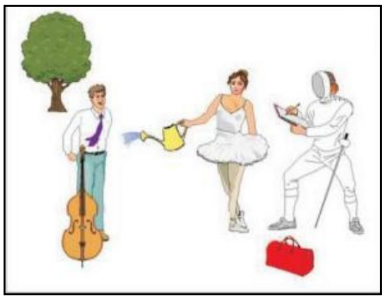

Figura 11 – Imagem utilizada no experimento de Knoerfele & Crocker (2006).

No segundo experimento, os participantes foram expostos a sentenças do alemão em que os elementos animados poderiam realizar ações estereotipadas (esperadas pelo ouvinte) ou inesperadas, tornando o elemento um agente único na compreensão da sentença. Por exemplo, espera-se de um mago que ele enfeitice alguém; de um detetive, que espie alguém; e de um piloto, que pilote um avião. Sentenças em que um detetive servia comida para alguém contrariariam a expectativa do ouvinte. Por exemplo, sentenças em que há um evento esperado como "*Den Piloten verzaubert gleich der Zauberer*" (O mago irá em breve enfeitiçar o piloto), e "*Den Piloten bespitzelt gleich der Detektive*" (O detetive irá em breve espiar o piloto) foram utilizadas. Sentenças do tipo "*Den Piloten bangadiert gleich der Zauberer*" (O mago irá em breve pôr um curativo no piloto) e "*Den Piloten verköstigt gleich der Detektiv*" (O detetive irá em breve servir comida ao piloto) exemplificam eventos inesperados.

Como resultados, os padrões de fixação ocular indicaram que "quando unicamente identificados como relevantes – tanto o conhecimento estereotipado quanto a informação relatada sobre quem faz o quê para quem possibilita a antecipação do agente apropriado na sentença"[5] (p. 512-513). As fixações mais frequentes no elemento que realizava a ação inesperada indicavam a relação entre as fixações oculares e a interpretação dos papeis temáticos da sentença ouvida. Ou seja, os participantes dedicaram mais tempo à identificação do elemento realizador da ação inesperada e posterior compreensão do evento.

De acordo com os autores, haveria duas etapas no processo de compreensão situada de sentenças:

> A CIA identifica dois estágios fundamentais na compreensão situada de sentenças. Primeiro, a compreensão de sentenças enquanto são ou-

[5] (…) when uniquely identified as relevant – both stereotypical knowledge and depicted information about who does what to whom enable anticipation of the appropriate agent role filler in the scene.

> vidas guia a atenção na cena, estabelecendo referência a objetos e eventos (Tanenhaus *et al.*, 1995; Knoerfele *et al.*, 2005), e antecipando referentes prováveis (Cf. Altman & Kamide, 1999). Uma vez que o enunciado tenha identificado o objeto ou evento mais provável, e a atenção tenha-se voltado para ele, a informação da cena analisada então rapidamente influencia a compreensão da sentença (...). Além disso, a CIA pressupõe que a relação temporal próxima entre a compreensão da sentença e a atenção à cena (Cf. Tanenhaus *et al.*, 1995) envolve uma estratégia de primeiro varrer a cena em vez de confiar somente no conhecimento linguístico/de mundo. Tal estratégia pode levar à preferência de elementos imediatamente representados ao conhecimento de eventos estereotipados na compreensão do que se observou no Experimento 2[6] (*Idem*, p. 523-524).

Os autores ressaltam que a capacidade de analisar elementos visuais imediatamente relacionados ao que se ouve é algo desenvolvido durante a aquisição de linguagem e, na fase adulta, o input linguístico ao qual um ouvinte/leitor é exposto cotidianamente geralmente não tem um correspondente visual imediato concomitantemente à exposição ao input linguístico. O processo de aquisição de linguagem, acreditam os autores, "pode ter, indubitavelmente, formado nossa arquitetura cognitiva, resultando em uma rápida interação entre sistemas cognitivos tais como a linguagem e a visão, e mecanismos de compreensão" (p. 526). Tais mecanismos, por sua vez, podem ter-nos permitido aproveitar a informação de uma cena rapidamente quando ela for relevante, especialmente em eventos da vida real, em que a informação sobre a cena é fugaz.

4.3
Estudos relacionando produção de linguagem e informação visual

Nesta seção, serão reportados trabalhos em produção da linguagem que exploram a questão da incrementalidade na passagem de informação do nível da conceptualização da mensagem para o nível da codificação gramatical, quando ocorre a formulação sintática da sentença (Griffin & Bock, 2000; Gleitman *et al.*, 2007; Myachykov, 2010). Esses trabalhos são relevantes para a presente pesquisa

[6] The CIA identifies two fundamental steps in situated utterance comprehension. First, comprehension of the unfolding utterance guides attention in the scene, establishing reference to objects and events (Tanenhaus *et al.*, 1995; Knoerfele *et al.*, 2005), and anticipating likely referents (see Altmann & Kamide, 1999). Once the utterance has identified the most likely object or event, and attention has shifted to it, the attended scene information then rapidly influences utterance comprehension (…). The CIA further assumes that the close time-lock between utterance comprehension and attention in the scene (e.g., Tanenhaus *et al.*, 1995) involves a strategy of first checking the scene rather than solely relying on linguistic/world knowledge. Such a strategy might lead to the greater relative priority of immediately depicted events over knowledge of stereotypical events in comprehension that we observed in Experiment 2.

tanto em termos teóricos quanto metodológicos. Do ponto de vista teórico, porque discutem se a formulação da sentença pode ocorrer antes de ter-se uma representação completa da ideia a ser expressa (no caso, uma proposição fechada), e em que medida propriedades de um dos argumentos de uma proposição (no caso de uma proposição que descreva uma cena com dois atores (um agente e um paciente, por exemplo) podem influenciar a seleção de uma dada estrutura sintática (por exemplo, ativa ou passiva). Ainda que esta dissertação tenha como foco a compreensão, interessa verificar o quanto a informação visual (no caso, foco atencional sobre um dos personagens de uma cena) poderia gerar expectativas em relação a estruturas sintáticas usadas para representar tal cena e também se uma dada estrutura sintática gera alguma expectativa sobre como uma ação com dois personagens será representada visualmente. Do ponto de vista metodológico, os trabalhos são relevantes pela utilização de um procedimento que permite dirigir o foco atencional do participante para elementos de uma cena. Esse recurso foi fundamental para que, nos experimentos aqui realizados, pudéssemos manipular experimentalmente o foco atencional dos participantes para elementos da cena que teriam, num plano abstrato/proposicional, papéis temáticos distintos.

4.3.1
O estudo de Griffin & Bock (2000): "What the eyes say about speaking"

Griffin & Bock (2000) realizaram estudos de produção de linguagem com monitoramento de movimentos oculares (*eye-tracking*) em que se pode verificar como o processamento de estímulo visual ocorre de maneira associada ao processamento e produção de linguagem. Nesse trabalho, sentenças da língua inglesa foram analisadas. Griffin & Bock (2000) basearam-se na teoria de produção de sentenças de Wundt (1900/1970), especialmente no argumento de que a produção de sentenças origina-se em uma conceituação holística e é seguida pela expressão sequencial de constituintes linguísticos. As autoras monitoraram os movimentos oculares dos participantes enquanto eles observavam desenhos em linha (*line drawings*). Quatro grupos de participantes foram formados. O primeiro grupo deveria descrever as figuras enquanto as via (discurso extemporâneo) e o segundo, após sua visualização (discurso preparado). Um terceiro grupo deveria realizar

uma tarefa não linguística de identificação do paciente. Por exemplo, para a figura abaixo, os participantes receberam a instrução de que deveriam identificar uma "vítima" (no caso, o rato que está sendo molhado pela tartaruga). O objetivo das autoras com essa tarefa era avaliar qual o tempo necessário para uma apreensão completa da cena.

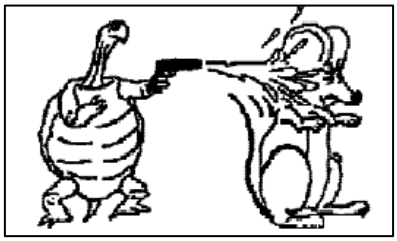

Figura 12 – Imagem utilizada no experimento de Griffin & Bock (2000).

Por fim, um quarto grupo de participantes via as figuras sem nenhuma tarefa específica. Com esse procedimento, pretendia-se verificar se havia algum elemento na cena que pudesse atrair a atenção do participante em especial, ou seja, se aspectos associados a particularidades das figuras poderiam torná-las mais salientes e, assim, influenciar os movimentos oculares dos participantes.

Na tarefa que envolvia apenas a inspeção da cena (quarto grupo acima indicado), não se observou preferência por nenhuma região particular da cena nos momentos iniciais de inspeção visual. Apenas após 1300ms, em média, do início da exibição da figura, uma região atraiu mais fixações que outras, tendo os pacientes recebido mais fixações que os agentes.

A comparação entre as tarefas de detecção do paciente e de produção extemporânea revelou que os participantes extraem rapidamente a estrutura de eventos representada nas figuras. Na tarefa de detecção, verificou-se que as fixações no paciente começam a divergir das fixações no agente aproximadamente 288ms após o início da apresentação da imagem e essa diferença atinge significância em 456ms. No caso da tarefa de produção extemporânea, a diferença entre agente e paciente começa aos 316 ms e atinge significância em 336ms. O tempo de resposta nas duas tarefas também foi bastante próximo: os participantes levaram 1690ms para indicar que haviam localizado o paciente e os falantes começaram a descrever oralmente a cena 1686 ms depois do início da apresentação da figura. No caso da tarefa de produção extemporânea, os dados indicam, ainda, que após a apreensão da cena, os movimentos oculares parecem ter sido guiados pelo processo de

formulação linguística; foi verificada uma relação em termos de ordem entre fixações oculares e ordem de palavras.

A comparação entre dados oculares da produção extemporânea e da produção preparada provê evidências adicionais para a hipótese de que ocorre apreensão da cena antes do início da formulação linguística. Assim como no discurso extemporâneo, na condição em que os participantes iniciam a formulação após a visualização da cena, diferenças entre fixações no agente e no paciente têm início após 304ms, atingindo significância em 472ms. Em ambos os grupos, a divergência entre fixações nos personagens da cena marca o início da atenção direcionada para a região correspondente à do elemento que vai figurar como sujeito da sentença.

Quanto ao processo propriamente dito de execução da fala, os resultados indicam que, no caso da fala extemporânea, o processo de seleção e de codificação fonológica dos nomes correspondentes ao sujeito e ao objeto ocorre de modo incremental. A fala do grupo que participou dessa condição foi menos fluente do que a do grupo da condição preparada. Essa redução de fluência sugere haver uma competição entre os processos de formulação e de execução na fala extemporânea.

Tomados em conjunto, os resultados obtidos nas quatro condições proveem evidências para a hipótese holística, na linha de Wundt, sobre a produção de sentenças. O processo de produção teria início com a apreensão ou geração de uma mensagem e prosseguiria com a formulação incremental de sentenças.

4.3.2
O estudo de Gletiman *et al.* (2007): "On the give and take between event apprehension and utterance formulation"

Gleitman e colaboradores (2007) realizaram dois experimentos investigando como a manipulação de atenção poderia afetar as escolhas dos observadores quanto a ordem de palavras, uso de verbos e estruturas sintáticas ao descrever imagens.

Os pesquisadores utilizaram metodologia similar àquela de Griffin & Bock (2000) e tentaram induzir, a partir de determinadas imagens e de um procedimento de manipulação de atenção visual, a produção de sentenças do tipo ativa, passiva,

com predicados em perspectiva (*perspective predicates*), predicados simétricos (*symmetrical predicates*) e sintagma nominal composto (*conjoined noun phrase*). A ideia de induzir a produção de tais estruturas foi motivada pelo fato de que estas codificam diferentes pontos de vista acerca de um evento. Na criação de cenas que pudessem eliciar a produção de verbos de perspectiva, havia imagens compatíveis com pares de verbos em inglês que poderiam ser classificados como de perspectiva, por exemplo, *buy/sell* (comprar/vender), *chase/flee* (perseguir/fugir), *win/lose* (perder/ganhar) e *give/receive* (dar/receber). Para os predicados simétricos em inglês, foram usadas cenas que pudessem ser descritas com verbos como *match*, *meet* e *argue*, os quais costumam ser empregados no plural e em estruturas intransitivas, indicando reciprocidade, como, por exemplo, em uma sentença do tipo "*The men met*" (Os homens encontraram-se). Para induzir a produção de sentenças com sintagma nominal composto, foram usadas cenas em que a ação estivesse sendo realizada por mais de um sujeito como as expressas por frases como "*The cat and the dog/The dog and the cat are growling at each other*" (O cão e o gato/O gato e o cão estão rosnando um para o outro). Havia ainda cenas cujas ações representadas poderiam ser descritas por verbos como "kick" (chutar), "scold" (repreender), "splash" (molhar), entre outros, a fim de eliciar sentenças do tipo "*The man is kicking the boy*" (O homem está chutando o menino) ou "*The boy is being kicked by the man*" (O menino está sendo chutado pelo homem). As imagens abaixo exemplificam alguns dos estímulos visuais utilizados no experimento de Gleitman *et al.* (2007):

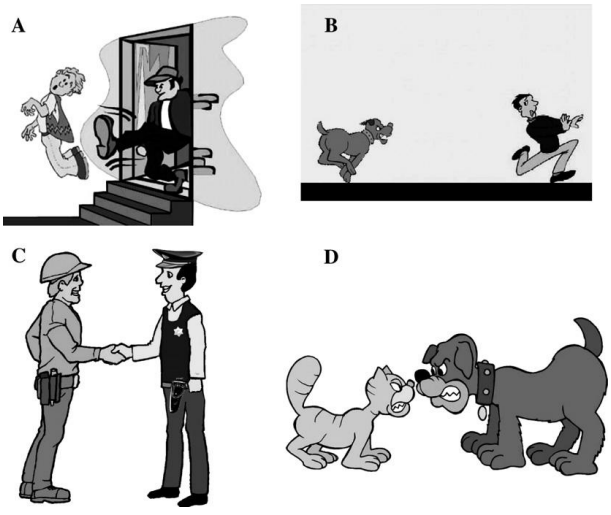

Figura 13 – Imagens utilizadas no experimento de Gleitman *et al.* (2007).

No experimento 1, o recurso de manipulação de atenção visual utilizado consistiu na apresentação de uma cruz de fixação ocular em um *slide* que aparecia antes do estímulo visual. Os participantes eram expostos a essa cruz de fixação ocular por 500 ms, depois viam o *slide* de manipulação da atenção por um período de tempo de 60 a 75 ms e, finalmente, o *slide* com a cena a ser descrita. Os pesquisadores relataram que, após a tarefa ter sido realizada, os participantes foram entrevistados sobre o experimento e nenhum declarou ter percebido que sua atenção visual havia sido manipulada.

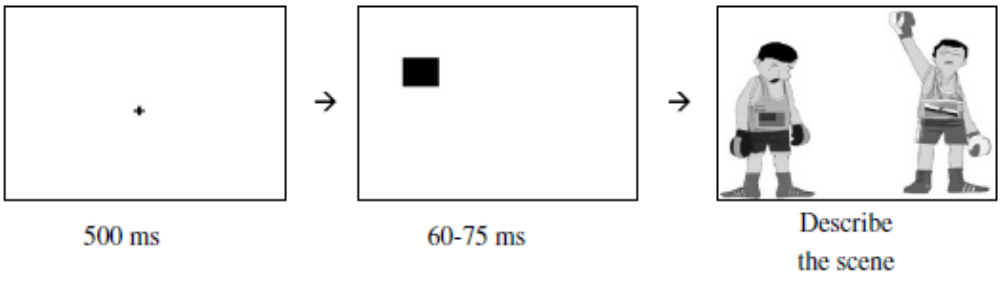

Figura 14 – Sequência de exposição da manipulação de atenção e cena com a instrução para descrição verbal utilizada no experimento de Gleitman *et al.* (2007).

A posição do recurso atencional – foco no agente ou no paciente, e orientação dos personagens – direita e esquerda foram manipuladas. As cenas usadas buscaram eliciar a produção de sentenças com verbos de perspectiva e sujeitos compostos.

As respostas dos participantes para os verbos de perspectiva foram submetidas à análise estatística por meio de teste ANOVA. Houve um efeito principal da pista atencional. Não houve também efeito de orientação esquerda-direita nem efeito de interação da localização da pista de captura de atenção com a orientação esquerda-direita dos personagens.

Em relação às sentenças do tipo sintagma nominal composto, os autores também realizaram uma ANOVA. Assim como nos verbos de perspectiva, houve efeito principal da pista atencional. Diferentemente dos verbos de perspectiva, houve efeito principal da orientação dos personagens – com tendência de se mencionar primeiro o personagem situado mais à esquerda na cena. Não houve, não obstante, interação significativa entre a localização da pista de atenção e a orientação esquerda-direita, ou seja, apesar da tendência acima reportada, a pista atencional teve efeitos equivalentes na definição do primeiro elemento mencionado,

independentemente de este estar posicionado à esquerda ou à direita na cena. Sintetizando, pode-se dizer que o principal resultado do experimento 1 foi que o personagem para o qual a atenção visual foi direcionada teve maior tendência de ser mencionado primeiro tanto para os verbos de perspectiva como para os sintagmas nominais compostos.

No Experimento 2, sentenças do tipo sintagma nominal composto foram utilizadas novamente, junto de sentenças ativas e passivas e predicados simétricos. As sentenças do tipo predicados em perspectiva foram eliciadas sem que houvesse manipulação de atenção na forma de quadrado preto sobre uma região da cena visual: as figuras foram apresentadas precedidas de uma cruz de fixação antes da figura, do mesmo modo que nas figuras distratoras. Nas figuras de sintagma nominal composto bem como nas que eliciavam a produção de sentenças ativas/passivas, a pista de manipulação de atenção foi posicionada à esquerda ou à direita de maneira equilibrada entre os estímulos.

A ANOVA realizada com os resultados da produção de sentenças do tipo sintagma nominal composto replicou os achados do experimento 1: verificou-se um efeito principal de localização da pista, bem como da variável orientação esquerda-direita. Não houve efeito de interação entre as duas variáveis.

Em relação à produção de sentenças com predicados simétricos, os resultados da ANOVA sugeriram efeito principal de localização da pista de atenção. Houve efeito marginal da variável orientação esquerda-direita, e não houve interação entre as duas variáveis analisadas.

Em relação às sentenças produzidas nas vozes ativa/passiva, a ANOVA indicou "um efeito pequeno, mas confiável de localização da pista de manipulação da atenção". A proporção média de sentenças em que o Agente foi mencionado foi de 0,85, ao passo que a proporção média de sentenças em que o Paciente foi mencionado foi de 0,74. Não houve efeito significativo de orientação esquerda-direita, nem interação entre orientação esquerda-direita e localização da pista de atenção. Os autores verificaram a diferença de tempo entre o início da produção de sentenças com o Paciente (sujeito não-preferido) e o início da produção de sentenças com o Agente (sujeito preferido): 2324 ms contra 2076 ms. Esse resultado

sugere que a produção de uma estrutura passiva requer maior tempo e esforço cognitivo do que a produção de uma sentença ativa.

De modo geral, os resultados acima reportados indicaram que há uma forte relação entre escolha da ordem dos referentes e os movimentos oculares: os personagens destacados pelas pistas de manipulação eram mais mencionados nas posições iniciais das sentenças. Note-se, contudo, que o efeito obtido ocorre muito antes do que foi verificado no experimento de Griffin & Bock (2000): os movimentos oculares para o agente e o paciente começaram a divergir durante os primeiros 200ms, resultado esse não esperado na hipótese holística, segundo a qual haveria um estágio inicial de extração/apreensão da essência da cena (*gist of the scene*) que não seria relacionado ao planejamento linguístico.

Quanto à produção de sentenças com verbos de perspectiva, embora de modo similar a Griffin & Bock (2000) não se tenha empregado recurso de manipulação de atenção, foi observado um efeito bastante inicial de escolha linguística: com maior número de olhares para o N1 do que para o N2 até mesmo nos primeiros 200ms (efeito esse significativo apenas na análise por participantes e não por itens).

Os autores concluem que "a apreensão e o planejamento linguístico sobrepõem-se temporariamente desde o início dos dois processos. Parece não haver um período de tempo em que os movimentos oculares são dissociáveis de fatores linguísticos durante as tarefas de descrição de cenas"[7] (*Idem*, p. 563). Acredita-se que os dois processos (visual e linguístico) ocorram em paralelo, tendo sua eficiência aumentada à medida que se acumula informação (*Idem*, p. 564). Outros fatores, como a ativação de lemas em decorrência da manipulação da atenção a um dos elementos visuais, e a hierarquização de elementos (por exemplo, no momento de visualização de uma cena de um cão perseguindo um homem), em que tanto o cão quanto o homem podem ser tomados como "figura" e o restante da cena como "fundo", oferecem "efeitos independentes mas adicionais na probabilidade de um elemento ser mencionado primeiro" na produção dos participantes (*Idem*, p. 565).

[7] (The results of Experiment 2) strongly support the claim that scene apprehension and linguistic planning temporally overlap from the onset of both processes. There does not appear to be a period of time in which eye movements are dissociable from linguistic factors during scene description tasks.

Finalmente, os autores concluem que, ainda que os participantes estivessem conscientes de que iriam descrever figuras e, por isso, seus sistemas visual, perceptual e linguístico pudessem estar previamente ativados, isso não pode querer dizer que sua capacidade de computação visual e linguística fosse ocorrer instantaneamente. A incrementalidade do processamento visual e linguístico ocorreria com eficiência à medida que tais sistemas operavam em paralelo e de maneira harmoniosa com o acúmulo de dados.

Em uma adaptação do experimento aplicado por Gleitman *et al.* (2007), Rodrigues & Barcellos (2013) realizaram um estudo de caráter exploratório, com falantes do português, com vistas a investigar o quanto a perspectiva a partir da qual se observa uma cena pode afetar as escolhas linguísticas do observador em sua descrição da mesma. Um experimento de produção induzida foi criado, focalizando o uso da voz verbal (ativa ou passiva) e verbos de perspectiva (perspectiva do agente/fonte ou do paciente/alvo, em verbos do tipo perseguir e fugir),

Os participantes foram expostos a cenas com manipulação de atenção sobre certos elementos visuais a fim de se verificar até que ponto a manipulação de atenção – um quadrado preto sobre um dos elementos visuais que aparecia antes de um dos referentes, agente ou paciente, durante 500 ms – afetaria a produção dos participantes. Por exemplo, se, em uma cena em que o foco atencional estivesse no agente, o participante produziria uma sentença na voz ativa e, inversamente, se uma sentença passiva seria produzida se o foco atencional estivesse no paciente. Das 16 sentenças experimentais, metade das cenas correspondentes teve manipulação de atenção no agente e metade, no paciente. A posição dos estímulos visuais (personagens) nas cenas foi controlada, com metade dos estímulos à direita e metade, à esquerda. Os participantes foram instruídos a produzir livremente sentenças descrevendo o que viam.

As análises estatísticas das respostas dos participantes indicaram que houve preferência por sentenças na voz ativa. Foi, contudo, identificada uma influência do recurso de manipulação de atenção quando este direcionava a atenção para o paciente. Nessa condição, foi verificada uma redução no número de estruturas ativas e um incremento na produção de passivas, em comparação à condição com foco no agente. Logo, embora possivelmente a escolha da estrutura sintática nes-

ses experimentos pareça ter sido determinada por questões de custo de processamento (ativas seriam menos custosas do que passivas), fatores de ordem atencional que colocam em perspectiva um dado elemento parecem também influenciar as decisões do produtor. Em que momento esses fatores atuariam é fruto de pesquisa que se encontra em desenvolvimento pelas autoras, a partir de experimentos em que fazem uso de rastreador ocular.

4.3.3
A relação entre a saliência de elementos do *display* e a escolha lexical na produção

Myachykov *et al.* (2011) realizaram um levantamento de estudos no qual discutiram o papel da atenção visual na produção de sentenças em que os observadores analisavam apenas dois personagens no *display*. Durante o período de aquisição de linguagem, costuma-se assumir que "quando se descreve um evento no mundo visual, os falantes tendem a formular suas sentenças conforme certa escolha estrutural e ordenamento linear de constituintes que se apresenta muito próximo da saliência visual dos protagonistas daquele evento" (Myachykov *et al.*, 2011, p. 96). Independente da língua usada para descrever o evento, a criação de uma mensagem *conceptual* (grifo dos autores) é necessária para compreendê-lo, codificando-o em um esquema temático (quem faz o quê para quem) (*Idem*, p. 97). Os autores iniciam esse levantamento com estudos com falantes de inglês, língua de ordem SVO, mais fixa, e, posteriormente, comentam sobre estudos com falantes de línguas de ordem mais flexível. Uma questão de fundo nessa comparação é em que medida escolhas eliciadas por pistas visuais diferem entre línguas, ou seja, apenas aspectos de ordem visual, como saliência perceptual, afetam as escolhas linguísticas ou é possível que aspectos da gramática das línguas possam vir a restringir o modo como a informação visual é descrita linguisticamente.

Pistas endógenas (oriundas das expectativas do falante) e exógenas (da saliência perceptual do referente) influenciam a seleção atencional do observador. Conforme já foi dito, os autores detiveram-se em experimentos em que apenas dois personagens estavam apresentados nos *displays*.

Os autores dividem esses experimentos de produção de linguagem em dois grupos: *priming*[8] referencial (*referential priming*) e *priming* perceptual (*perceptual priming*). Nos experimentos de *priming* referencial, o participante pode observar um dos referentes envolvidos na cena antes que a cena apareça por completo na tela. Estudos realizados em psicolinguística nos anos 1960 pesquisados pelos autores sugeriram que informações visuais, semânticas e lexicais associadas ao referente visualizado primeiro facilitarão a produção e/ou a compreensão de sentenças que o envolvam (*Idem*, p. 98). Os autores citam um experimento realizado por Prentice (1967), com tarefas de verificação de sentenças, em que o referente visualizado primeiro era rapidamente identificado por um observador quando a sentença-alvo começava com o elemento pré-visualizado ou quando ele estivesse na posição de sujeito. Quando tinham de descrever eventos transitivos simples e o elemento pré-visualizado era o agente, os participantes tendiam a produzir sentenças na voz ativa, por exemplo, "*The fireman is kicking the cat*" (O bombeiro está chutando o gato); quando o elemento pré-visualizado era o paciente, os participantes tendiam a produzir sentenças na voz passiva, por exemplo, "*The cat is being kicked by the fireman*" (O gato está sendo chutado pelo bombeiro). Esses estudos influenciavam os participantes a produzirem sentenças com o elemento destacado em posição inicial; todavia, esse efeito não era tão forte: enquanto os testes em que o referente tomado como agente eliciavam 100% de sentenças na voz ativa, os testes em que o referente tomado como paciente eliciavam 40-50% de sentenças na voz passiva. Além disso, em tarefas de verificação de sentenças com *priming* referencial, os participantes eram mais rápidos na verificação de sentenças na voz ativa que de sentenças na voz passiva. Os autores mencionam o estudo de Clark & Chase (1972, Cf. item 4.4), de verificação de sentenças, cujos resultados sugeriam verificação mais rápida e precisa quando as sentenças começavam com o referente destacado na posição de sujeito (*Idem*, p. 99).

Os estudos de *priming* referencial não envolveriam somente manipulação da atenção visual, pois o observador é exposto a informações perceptuais e semânticas sobre o referente. Os estudos de *priming* perceptual, que constituem uma va-

[8] *Priming* é um aumento na velocidade de reconhecimento de uma palavra por influência de exposição anterior a outra palavra com a qual se mantenha proximidade. Field (2004) cita um exemplo em que, uma vez exposto à palavra "doutor", um participante reconhecerá "enfermeira" ou "paciente" mais rápido. A palavra "doutor" seria o prime e "paciente", o alvo. Costuma-se dizer que o item lexical "doutor" exerceria *priming* sobre o item lexical "paciente".

riante do paradigma da pista visual (Posner, 1980), caracterizam-se por manipulação da atenção visual sobre um dos referentes e buscam verificar em que medida isso pode influenciar a produção do observador.

Myachykov e colaboradores fazem menção ao experimento de Tomlin (1995, 1997) que, conforme foi visto no capítulo 2, ficou conhecido por salientar referentes a fim de induzir a produção de sentenças na voz ativa e na voz passiva. Os autores fazem referência também ao estudo de Gleitman *et al.* (2007), reportado na seção 4.3.2, que confirmou os resultados de Tomlin, mas com efeitos mais fracos devido, provavelmente, à manipulação de atenção mais sutil (um quadrado de manipulação de atenção aparecia 65 ms antes do referente).

De acordo com Myachykov *et al.* (2011, p. 101), dados obtidos apenas com falantes de inglês não permitem estabelecer uma diferença entre um ordenamento linear e um ordenamento gramatical (por papeis temáticos) das sentenças. No inglês, o papel de sujeito tende a ser confundido com o elemento em posição inicial. Os autores acreditam que o efeito de *priming* perceptual aconteça na seleção estrutural de sentenças do inglês. Isoladamente, a saliência perceptual do referente já pode predizer uma série de escolhas de seleção estrutural, além disso, tal efeito depende da associação entre ele e o referente.

Estudos com falantes de línguas de ordens mais flexíveis que o inglês (Myachykov *et al.*, 2010), que analisaram efeitos de *priming* perceptual ao se compararem sentenças do finlandês e do inglês, bem como sentenças do russo (Kaiser & Vihman, 2006) e do coreano (Hwang & Kaiser, 2009), indicaram que a manipulação da atenção visual teve algum efeito nas fixações iniciais, no entanto, não indicou efeito de *priming* perceptual a ponto de afetar a escolha lexical dos participantes. Myachykov e colaboradores acreditam que "os falantes universalmente tentam empregar o mecanismo de papeis temáticos para representar o elemento salientado no plano estrutural da sentença" (2011, p. 103). Parece haver uma tendência entre os falantes a associar o referente destacado pela manipulação ao papel de sujeito sentencial. Outras ordens de palavras assumiriam um papel secundário, sendo empregadas quando não se puder atribuir o papel de sujeito ao elemento destacado diretamente. Na interpretação dos autores, a manipulação de atenção visual não teve tanta influência quanto a tendência dos observadores a determinar, inicialmente, um sujeito realizador da ação apresentada nos estímulos visuais. Ou seja, parece que, ao se expor um observador a uma cena, é possível

que sua busca inicial seja orientada pela busca de um elemento que corresponda ao agente da ação.

4.4
Uma proposta de integração entre informação linguística e visual (Clark & Chase, 1972)

Clark & Chase (1972) desenvolveram quatro experimentos de comparação entre sentenças e imagens. Os autores partiram de certas considerações gerais para estabelecer uma teoria da comparação sentença-imagem. Segundo eles, "para que uma sentença e uma imagem sejam comparadas, elas devem ser representadas, fundamentalmente, no mesmo formato[9]" (p. 473), o que vem ao encontro da ideia apresentada neste trabalho. Os autores defendem que "o formato mental codifica as interpretações em vez das propriedades perceptuais de sentenças e figuras"[10] (p. 473).

Nos quatro experimentos, os participantes tinham que ver um *display* ora contendo uma sentença e uma figura posicionados em lados diferentes do mesmo, ora *displays* separados contendo sentenças e figuras. Os participantes foram expostos a estímulos em que liam, à esquerda do *display*, uma sentença como "The star isn't below the line" (A estrela não está abaixo da linha) e "The star is above plus" (A estrela está acima do sinal de mais) e, à direita do *display*, visualizavam uma figura de uma estrela acima de uma linha, tendo, em seguida, que pressionar botões em um dispositivo que indicavam o julgamento de verdade verdadeiro ou falso. As sentenças sempre apresentavam as palavras "above" (acima) e "below" (abaixo) e descreviam a localização vertical de uma das figuras. Por sua vez, as figuras geométricas utilizadas estavam sempre posicionadas uma acima da outra.

Os autores adotaram uma abordagem procedimental de como ocorreria o processamento de tais dados. Primeiro, o sujeito realizaria a representação mental de sentenças locativas, em que uma sentença do tipo "A está acima de B" seria similar a $(A\ acima\ de\ B)_{Sentença}$. Em seguida, ocorreria a representação mental da figura, sendo sucedida pela comparação das representações entre si e, finalmente, ocorreria a produção da resposta, isto é, o participante realizaria a tarefa propria-

[9] (…) for a sentence and picture to be compared they must be represented, ultimately, in the same format.
[10] (…) the mental format codes the interpretations rather than the perceptual properties of sentences and pictures.

mente dita: os participantes fariam um julgamento de verdade entre a possível correspondência entre sentença e figura. Como resultados, os autores verificaram uma correspondência entre os resultados e modelos de julgamentos de verdade que haviam sido desenvolvidos previamente e aqueles obtidos com os participantes.

Os pesquisadores consideraram que, no estágio 2 do processamento dos dados, "a figura seria codificada de modo a ser comparada com a representação da sentença no estágio 3" (p. 510). À época da produção do estudo, os autores não tinham muitas informações sobre como caracterizar as representações das sentenças e das imagens, no entanto, acreditavam que havia "um sistema 'interpretativo' comum que deve ser acionado por um conjunto de princípios independentemente da origem linguística ou perceptual de uma dada interpretação"[11] (p. 515).

A proposta de Clark & Chase (1972) é compatível com a proposta de Pylyshyn (1978, Cf. item 2.3.2), no que concerne à ideia de que o resultado do processamento do input visual é um conteúdo proposicional. Além disso, ainda que Clark & Chase (1972) não se refiram a módulos responsáveis pelo processamento exclusivo de dados de ordem linguística e/ou visual, sua proposta aproxima-se da ideia aqui defendida de que a teoria modular (na linha de Fodor, 1983) poderia explicar o processamento de informações de diferentes ordens. Cumpre notar, contudo, que, como salienta Anderson (1976), os pesquisadores não explicitaram claramente quais princípios norteariam a passagem de uma sentença à sua representação proposicional; não se aborda a especificidade de cada tipo de sentença investigada no que tange à sua representação, de modo que seu modelo não é falsificável de maneira tão simples (p. 29).

No próximo capítulo, apresentaremos os experimentos construídos para fins desta dissertação. Conforme será visto, os experimentos fazem uso de estímulos linguísticos combinados a estímulos visuais de modo a investigar como se dá a interação entre informação proveniente dos dois módulos cognitivos. Nesse sentido, os trabalhos reportados neste capítulo, embora com focos distintos e/ou voltados para outra modalidade (produção da linguagem), proveem elementos tanto de ordem teórica quanto metodológica para a presente pesquisa.

[11] Underlying both language and perception, we have argued, is a common "interpretive" system that must be handled by one set of principles no matter whether the source of a particular interpretation is linguistic or perceptual.