

6

Considerações finais

Esta pesquisa procurou discutir aspectos relacionados à integração entre informação linguística e visual em situações experimentais que envolvem comparação de sentenças com imagens. Como visto na introdução, apesar de técnicas que fazem uso de imagens serem bastante utilizadas em pesquisas em processamento linguístico, pouco se sabe acerca de como material visual é percebido e processado e são raras as pesquisas que se ocupam diretamente da interface linguagem/visão. É nesse hiato que o presente trabalho se insere.

Examinou-se especificamente em que medida informações de natureza linguística podem antecipar ou gerar expectativas quanto à organização da informação visual, assim como em que medida a organização da informação visual pode gerar expectativas sobre as estruturas linguísticas. Duas hipóteses de trabalho orientaram os experimentos desenvolvidos para este trabalho.

A primeira hipótese de trabalho relaciona a direcionalidade da leitura da sentença (da esquerda para a direita) e a posição dos elementos correspondentes ao sujeito das sentenças. Acreditava-se que, em tarefa de verificação de congruência entre sentença e imagem, o curso da leitura (da esquerda para a direita) direcionaria o curso do mapeamento visual, gerando uma expectativa de que o elemento correspondente ao sujeito da sentença estivesse também posicionado à esquerda da imagem. Por outro lado, esperava-se que o elemento correspondente ao sujeito sentencial, uma vez posicionado à direita da imagem, viesse a gerar maior dificuldade para ser identificado.

A segunda hipótese de trabalho diz respeito à relação entre, de um lado, papéis temáticos correspondentes aos elementos/personagens envolvidos em uma dada cena e, de outro, o *tipo de estrutura* sintática e a escolha lexical usados para descrevê-la. A hipótese é a de que o foco atencional num dado elemento da cena, com um determinado papel temático (agente/fonte ou paciente/alvo), geraria expectativas em relação ao tipo de sentença/escolha lexical apresentados para descrever a cena.

Para realizar o estudo, foi necessário conjugar referências bibliográficas tanto da área de processamento linguístico quanto de processamento visual. Os capí-

tulos de 2 a 4 cobrem a resenha da literatura. No capítulo 2, foram comentados os aspectos biológicos da visão (desde a trajetória da luz no olho até áreas do córtex cerebral responsáveis pelo processamento visual) assim como teorias sobre o reconhecimento de objetos. Em seguida, uma discussão sobre a visão como sistema de processamento de informação foi apresentada, tomando como referência trabalhos de Henderson (2003), Marr (1982) e Biederman (1987). A partir do texto de Henderson (2003), considerou-se como processos *bottom-up* e *top-down* interagiriam no processamento visual de imagens. Foi visto que, enquanto Marr (1982) e Biederman (1987) detiveram-se mais especificamente na descrição de que algoritmos possibilitariam o processamento de informação visual, sem considerar outras modalidades, Paivio, na década de 1960, propôs a teoria da dupla codificação, em que apresentou a ideia de processamento multimodal (Paivio 1991). A esse autor, contrapôs-se a teoria proposicional de Pylyshyn (1978), segundo a qual as representações mentais não seriam registradas no cérebro em uma relação de semelhança com a imagem, mas sim de forma mais abstrata (Sternberg, 2008, p. 229). Ainda nesse capítulo, dada a relevância de processos atencionais para os experimentos realizados para esta pesquisa, realizou-se uma breve apresentação, em uma perspectiva da psicologia cognitiva, do trabalho de Posner (1980), dado seu pioneirismo em investigar a influência de pistas visuais na execução de tarefas, e do experimento de Tomlin (1997), com pistas visuais explícitas.

No capítulo 3, enfocou-se a questão da modularidade da mente, buscando situar, nesse contexto teórico, como se poderia caracterizar a relação entre a linguagem e outros módulos da cognição. Para isso, abordou-se, inicialmente, a teoria da modularidade da mente de Fodor (1983) e a proposta da modularidade representacional de Jackendoff (1996, 2002), que se ocupa, entre outros aspectos, da comunicação intermodular. A teoria da modularidade representacional centra-se na ideia de que a mente codifica dados em formatos oriundos de módulos específicos, contudo, diferentemente de Fodor (1983), o autor propõe que existiriam módulos híbridos, cujo conteúdo representacional seria função dos módulos em comunicação. Em seguida, apresentou-se a caracterização proposta por Hauser, Chomsky & Fitch (2002) sobre a faculdade da linguagem, a qual contribui para esta pesquisa por indicar como essa faculdade, em um sentido amplo, interagiria com os demais sistemas cognitivos da mente, em especial o sistema conceitual-

intencional e o sistema articulatório-perceptual. A posição defendida neste trabalho é compatível com essa visão da faculdade da linguagem no sentido amplo. Considera-se que o material linguístico enviado, via forma lógica, para o sistema conceptual-intencional, poderia apresentar nesse sistema um formato proposicional, semelhante ao das representações provenientes do módulo visual conforme a proposta de Pylyshyn (1978). Dessa forma, a interface entre os módulos linguístico e visual não exigiria um módulo híbrido, como na proposta de Jackendoff; a interface poderia ocorrer com base em representações proposicionais no sistema conceptual-intencional, resultantes do processamento de informações nos módulos linguístico e visual.

No capítulo 4, apresentaram-se estudos de compreensão e de produção de linguagem que utilizaram imagens. Inicialmente, mencionaram-se estudos inseridos no paradigma do mundo visual (*visual world paradigm*), que teve sua origem em estudos de Cooper (1974) e ficou popularizado por testes elaborados por Tannenhaus *et al.* (1995). Tal metodologia analisa padrões de movimento ocular e fixações conforme instruções linguísticas e influenciou investigações sobre processos preditivos e incrementalidade (Altmann & Kamide, 1999; Kamide *et al.*, 2003; Knoerfele & Crocker, 2006). Também foi comentada a teoria de Knoerfele & Crocker (2006) da Inter-Relação Coordenada (*Coordinated Interplay Account*), que defende uma interface entre dados de ordem linguística e visual considerando as dimensões temporal (*temporal dimension*), e informacional (*informational dimension*) em um processo incremental de compreensão de sentenças.

Ainda nesse capítulo, são comentados dois estudos relacionando produção de linguagem e informação visual, os quais nortearam a metodologia dos experimentos realizados nesta pesquisa – as investigações de Griffin & Bock (2000) e Gleitman *et al.* (2007), ambos estudos com falantes nativos de inglês. O trabalho de Griffin & Bock (2000) partiu da investigação do ponto inicial da produção linguística dos participantes em uma tarefa de descrição de uma cena. As autoras, a partir da relação entre elementos fixados inicialmente em uma cena e tipos de estruturas linguísticas usadas para descrevê-la, chegaram à conclusão que a codificação gramatical só teria início após o processo de formulação de uma mensagem completa. Verificaram que, quando os participantes iniciam a formulação de sentenças, já teriam realizado o mapeamento geral da cena (*gist of the scene*).

O trabalho de Gleitman *et al.* (2007) aprofunda a investigação de Griffin & Bock (2000), utilizando manipulação de atenção em um dos elementos visuais do *display*. Nesse estudo, foram usadas cenas que pudessem eliciar a produção de sentenças do tipo ativa e/ou passiva, verbos de perspectiva (*perspective verbs*), verbos simétricos (pronominais), bem como sentenças de sujeito composto. As cenas que eliciavam sentenças do tipo ativa ou passiva sofreram manipulação de atenção, o que parece ter afetado a ordem de palavras na produção dos participantes. O elemento destacado pela manipulação de atenção tendia a ser mencionado primeiro. As cenas que eliciavam sentenças com verbos de perspectiva não sofreram manipulação de atenção; no entanto, as fixações iniciais pareciam determinar o início da produção linguística dos participantes. Os resultados obtidos pelos autores são compatíveis com a hipótese de que o processamento linguístico na passagem da conceptualização da mensagem para a codificação linguística seria incremental.

Esses trabalhos contribuíram para a definição da metodologia dos experimentos realizados, tanto quanto ao tipo de informação linguística analisada – sentenças dos tipos ativa e passiva e verbos de perspectiva, quanto ao tipo de estímulo (desenhos e fotografias relacionados às sentenças experimentais) e ao emprego de recurso de manipulação de atenção. Efeitos de orientação esquerda/direita do input visual que foram analisados naqueles trabalhos também foram investigados nos experimentos aqui realizados, assim como a focalização atencional ora no personagem agente, ora no paciente, nos *displays*. Também foram comentadas questões de saliência perceptual no *display* e sua influência na escolha lexical dos participantes no momento da produção de linguagem, investigadas por Myachykov *et al.*, (2011). Esse trabalho discutiu a inter-relação entre pistas endógenas (expectativas do falante sobre os estímulos linguístico/visual) e exógenas (saliência perceptual de um referente na cena).

Como os experimentos realizados para este trabalho envolveram tarefas de verificação de sentença (*sentence verification tasks*), mencionou-se, ao final do capítulo 4, a pesquisa de Clark & Chase (1972), por terem sido um dos primeiros a investigar a integração entre informação de ordem linguística e visual, buscando caracterizar o tipo de representação que permitiria essa integração.

No capítulo 5, os dois experimentos realizados para investigar a relação linguagem-visão são reportados. Antes, no entanto, comentou-se sobre a técnica de comparação sentença-figura (Levelt, 1978), referência para a metodologia dos experimentos realizados para este trabalho.

Os experimentos tiveram como objetivos específicos verificar, principalmente com base nas médias de *tempo de resposta* dos participantes, (i) relação entre direcionalidade de leitura de uma sentença e o mapeamento visual da cena, com foco no personagem correspondente ao sujeito da sentença; (ii) se o tipo de sentença (ativa ou passiva) e de verbo (perspectiva) podem influenciar o mapeamento de uma cena em relação à localização do elemento agente e do paciente, e (iii) se a manipulação de atenção visual em um dos personagens (agente/fonte ou paciente/alvo) pode contribuir para a antecipação da estrutura da sentença que a descreve.

No **primeiro experimento**, buscou-se verificar se a posição do elemento visual correspondente ao sujeito da sentença em uma fotografia (esquerda vs. direita) iria interferir na verificação da sentença correspondente. Duas situações experimentais foram elaboradas de modo a avaliar efeitos de ordem na apresentação dos estímulos visual e linguístico. Na primeira, os participantes leram a sentença antes de verem a fotografia; na segunda, o oposto ocorreu.

Observando os resultados da análise de variância do grupo exposto à ordem de estímulos sentença/fotografia para as variáveis *tipo de sentença* e *posição do sujeito*, verificou-se que não houve efeito principal dessas variáveis no comportamento dos participantes; houve, contudo, efeito de interação, tendo sido a condição passiva direita mais custosa para o processamento. Na análise de variância para as condições experimentais envolvendo *expressão de perspectiva do verbo* e *posição do sujeito*, o tipo de expressão (neutra, perspectiva 1 ou perspectiva 2) interferiu no comportamento dos participantes. Também foi observado efeito de interação entre as variáveis. Em relação ao efeito principal, parece haver uma escala de custo de processamento – neutra < perspectiva 1 < perspectiva 2. As maiores médias de *tempo de resposta* dos verbos de perspectiva 2 podem estar relacionadas a questões de acesso lexical: o verbo de perspectiva 2, que empreende um sujeito paciente/alvo, no momento da compreensão, precisaria ser acessado junto

com seu par lexical (comprar/vender, por exemplo). Quanto à interação, o mais relevante é o contraste entre perspectiva 2 esquerda vs. direita, sendo esta última condição a mais custosa de todas.

A análise de variância dos dados do grupo exposto à ordem de estímulos fotografia/sentença para as variáveis *tipo de sentença* e *posição do sujeito* indicou efeito principal de *tipo de sentença*, bem como efeito de interação entre *tipo de sentença* e *posição do sujeito*. Assim como na primeira situação, as ativas foram processadas mais rapidamente do que as passivas. Em relação a esse efeito de interação, foi visto que alguns itens experimentais na condição ativa/sujeito à esquerda podem ter gerado uma interpretação distinta da expressa pela sentença usada para descrever a cena, o que deve ter elevado seus tempos de resposta, contrariamente às previsões da hipótese 1. A análise de variância para as variáveis *expressão de perspectiva pelo verbo* e *posição do sujeito* indicou que houve efeito principal e de interação das duas variáveis. O problema com as imagens acima comentado também impactou os tempos médios de resposta nessa segunda análise, pois a condição ativa/sujeito à esquerda correspondia à condição neutra/sujeito à esquerda; logo o resultado relativo a efeito principal de posição precisa ser tratado com cautela. Possivelmente, na ordem sentença/fotografia, não houve interferência dos estímulos visuais citados, porque, como a sentença precedia a fotografia, o conteúdo expresso já deve ter limitado a interpretação da imagem. Quanto aos níveis da variável *expressão de perspectiva pelo verbo*, assim como na primeira situação experimental, perspectiva 2 foi a condição mais custosa. Não houve, não obstante, diferença entre perspectiva 1 e neutra, possivelmente porque a condição neutra teve um tempo maior de leitura decorrente de um item experimental que obteve alto índice de respostas “não”. Quanto ao efeito de interação, destacam-se os maiores tempos de verbos de perspectiva 2 quando o elemento visual correspondente ao sujeito estava posicionado à direita do *display*, corroborando previsões e indicando, dessa forma, maior complexidade de processamento.

Com o objetivo de analisar a influência da *ordem dos estímulos* no comportamento dos participantes, realizou-se uma análise de variância tomando **situação experimental como fator grupal**. O importante a destacar aqui é que não houve efeito principal dessa variável em nenhuma das análises conduzidas.

No **segundo experimento**, buscou-se verificar se a manipulação de atenção sobre o elemento agente/fonte ou paciente/alvo teria influência no comportamento dos participantes durante a tarefa. **Não** se obteve efeito principal de foco atencional tanto ao analisar a variável *tipo de sentença* quanto ao analisar *expressão de perspectiva pelo verbo*. Possivelmente neste segundo experimento, em que o participante primeiro via a figura, deve ter ocorrido um mapeamento completo da cena, que, de algum modo, obscureceu/neutralizou o efeito que se esperava do recurso de manipulação de atenção, que pretendia colocar em evidência um dos personagens. Em pesquisas futuras, seria interessante investigar uma ordenação dos estímulos em que a sentença preceda a figura fazendo uso de equipamento de *eyetracker*. O único efeito obtido nesse experimento foi para tipo de sentença, com passivas com maior tempo de resposta do que ativas, e para *expressão de perspectiva pelo verbo*, reproduzindo a escala de complexidade já observada no primeiro experimento – neutra < perspectiva 1 < perspectiva 2.

Analisando os dados dos dois experimentos, à exceção de uma das situações experimentais com fotografias, da mesma maneira como no experimento realizado por Gleitman *et al.* (2007), não houve efeito de *posição do sujeito (left-right orientation)*. Cumpre lembrar que os autores não utilizaram manipulação da atenção para as sentenças com *expressão de perspectiva pelo verbo*. Por outro lado, o monitoramento ocular registrado por Gleitman *et al.* durante os testes indicou que os personagens visualizados primeiro pelos participantes tendiam a ser mencionados primeiro e isso influenciava sua escolha lexical (qual verbo de perspectiva usar para descrever a cena). Gleitman *et al.* (2007, p. 564) mencionam a preferência dos participantes por verbos como *chase* (perseguir) em vez de *flee* (fugir). Em nosso experimento, isso pode estar refletido nas médias de *tempo de resposta* dos verbos de perspectiva 1 (verbos que empreendem sujeito agente/fonte), inferiores aos verbos de perspectiva 2 (verbos que empreendem sujeito paciente/alvo).

Uma questão interessante para reflexão é a da definição de verbos de perspectiva no português, como classificar seus argumentos e como eles são representados no léxico mental, além de definição de papéis temáticos para cada tipo de verbo. Na elaboração dos experimentos, utilizou-se os termos “agente/alvo” e “paciente/fonte” como artifício para referir-se a um grupo de verbos de complexa

classificação argumental, mas sabemos que não é livre de problemas/dificuldades essa classificação.

Os resultados obtidos a partir dos experimentos realizados mostram-se compatíveis com parte considerável da literatura psicolinguística e poderão ser úteis para a caracterização de processos de conceptualização de informações linguísticas em falantes do português quando expostos a informação visual. Como desdobramento deste trabalho, testes com uso de equipamento de monitoramento ocular e com estímulo linguístico sonoro concomitante à exposição do observador à imagem poderiam ser feitos. Esse tipo de pesquisa pode vir a prover dados sobre como se dá o processamento incremental de informação linguística e visual e sobre como as representações provenientes dessas modalidades são integradas.

Também merecem ser investigados, em pesquisas futuras, fatores associados à saliência visual e estímulos buscando eliciar outras estruturas, como predicados simétricos (A mulher e o homem se beijaram/ se cumprimentaram, etc.) com diferenças de posicionamento na imagem. Outra questão relevante seria contrastar dados do português com os de outras línguas, com ordem mais livre, de modo a buscar avaliar em que medida restrições de ordem gramatical de cada língua poderiam influenciar nos resultados.