

PROCESSAMENTO INCREMENTAL DE SENTENÇAS E PROCESSOS DE PERCEÇÃO VISUAL: QUESTÕES TEÓRICAS E METODOLÓGICAS

Aluna: Jessica Silva Barcellos
Orientadora: Erica dos Santos Rodrigues

Introdução

Este relatório tem por objetivo reportar as atividades desenvolvidas nos primeiro ano do projeto de iniciação científica voltado à investigação da interação entre informação linguística e visual no processamento incremental de sentenças. A pesquisa vincula-se diretamente ao projeto de RODRIGUES [1] e é conduzido no âmbito do LAPAL (Laboratório de Psicolinguística e Aquisição da Linguagem – PUC-Rio).

Conforme apontam Ferreira & Tanenhaus[2], a despeito de se voltarem para problemas semelhantes e fazerem uso de vocabulário similar na investigação de vários fenômenos, pesquisas relativas a processamento linguístico e processamento visual têm sido conduzidas, em grande parte, de modo independente, especialmente quando se considera a língua falada. Do ponto de vista metodológico, observa-se uma aproximação entre as duas áreas: são considerados como variáveis dependentes em experimentos sobre percepção visual e processamento linguístico, respectivamente, respostas linguísticas a partir de estímulos visuais e o desempenho em tarefas visuais a partir de comandos linguísticos. No entanto, na condução de experimentos pelas duas áreas, pouco se considera das pesquisas desenvolvidas em separado. Em experimentos em que se busca avaliar percepção visual, por exemplo, tarefas de nomeação são empregadas, mas sem levar em consideração questões ligadas à representação e acesso lexical. Nos testes psicolinguísticos, por sua vez, não é incomum o emprego de imagens sem que se observem determinadas propriedades do estímulo visual bem como o momento de apresentação de informação linguística e visual (se concomitante ou não).

Em nosso projeto, buscamos aprofundar o conhecimento acerca das propriedades dos estímulos visuais que podem ser tomados como relevantes em situações experimentais que fazem uso de imagens e também avaliar questões relativas ao curso temporal da produção de sentenças a partir de experimentos que explorem a interface linguagem/visão.

Entre os objetivos da pesquisa, em seu primeiro ano, elencamos os seguintes: (i) caracterizar, com base em bibliografia na área de percepção visual, propriedades do estímulo visual consideradas relevantes para a análise de objetos e cenas; (ii) especificar os tipos de imagens empregados em experimentos psicolinguísticos de modo a identificar propriedades dos estímulos visuais que possam, potencialmente, influenciar respostas dos sujeitos; (iii) investigar a partir de experimentos com imagens questões ligadas a como se dá a passagem de um nível de conceptualização da mensagem para a formulação propriamente linguística de enunciados.

Em relação especificamente ao objetivo (iii), buscou-se analisar em que medida a perspectiva a partir da qual uma cena/imagem é considerada pode afetar as escolhas linguísticas na descrição da referida cena e quão incremental seria esse processo. A ideia de incrementalidade aponta para o caráter gradual do processamento da fala, ou seja, para o fato de o planejamento da mensagem não precisar estar completamente concluído para que o falante inicie a produção. Levelt e colaboradores [3] consideram a existência de diferentes níveis/etapas no processo de produção de fala: conceptualização

da mensagem, codificação gramatical, codificação fonológica e articulação. A passagem do primeiro nível para o segundo é motivo de discussão e pode suscitar a seguinte questão: a representação conceptual precisa ser finalizada para que se dê início à codificação gramatical ou, tão logo a conceptualização é iniciada, informações de natureza gramatical também começam a ser processadas.

Nesta pesquisa, buscamos, através de metodologia experimental, investigar se a apreensão de uma imagem e o início da produção da fala ocorrem em paralelo. Para isso, foram construídos dois experimentos de produção induzida de descrição de imagens. No primeiro deles, focalizou-se o emprego da voz verbal (estruturas ativas e passivas) e de determinados tipos de verbos que codificam perspectiva (perspectiva do agente/fonte ou do paciente/alvo – exemplo: perseguir vs. fugir). Já no segundo, voltou-se especificamente para o emprego da voz verbal. Neste relatório, resultados do primeiro experimento e resultados parciais do segundo são apresentados.

Em termos de sua organização, o texto está dividido em três seções: resenha da literatura, metodologia e conclusões.

Resenha da Literatura

Previamente à montagem dos experimentos, foram realizadas leituras na área de percepção visual acerca de propriedades do estímulo visual relevantes para a análise de objetos e cenas; foram analisados tipos de imagens empregados em experimentos psicolinguísticos de modo a identificar propriedades de ordem *bottom-up* que possam, potencialmente, influenciar respostas dos sujeitos em testes linguísticos.

Apesar de o número de pesquisas que exploram a interface linguagem/visão crescer a cada dia, muitas delas ainda são baseadas em questões que não foram muito bem definidas, como por exemplo: o que são cenas, como elas são processadas e como os movimentos oculares são direcionados em uma cena.

Henderson e Ferreira [4] compilaram estudos sobre percepção de cenas naturais, concentrando-se principalmente em como cenas podem ser definidas e como elas são reconhecidas. Os estudos sobre percepção focalizam-se em cenas do mundo real ou naturais, uma vez que esses estão interessados no processamento de estímulos visuais passíveis de ocorrer no dia a dia. Entretanto, essa mesma ideia não é aplicada em testes psicolinguísticos que investigam essa questão. Nos experimentos, os estímulos costumam ser desenhos de linha, desenhos mais elaborados, fotografias, desenhos reproduzidos a partir de fotografias, um conjunto de objetos dispostos espacialmente a fim de representar uma cena real, ou seja, representações do real, mas raramente o real em si. Isso certamente influencia nos resultados, uma vez que cada estímulo possui suas propriedades específicas, como preenchimento do campo visual, intensidade e regularidades espaciais, podendo direcionar os movimentos oculares para diferentes regiões. Em experimentos psicolinguísticos que utilizaram cenas do mundo real (*real world scenes*) como estímulos visuais, participantes foram observados durante tarefas cotidianas, como fazer um chá, preparar um sanduíche, lavar as mãos e dirigir. Foi verificado que os movimentos oculares dos participantes diferiam dos registrados em experimentos com representações visuais (*depictions*). Um exemplo de parâmetro para o qual foram verificadas diferenças entre os dois tipos de estudo é a amplitude média da sacada. Enquanto em testes com representações a amplitude dos movimentos sacádicos ficou abaixo de um grau, sendo raras sacadas com dez graus, nos testes com cenas do mundo real a amplitude média aproximou-se de cinco graus e sacadas com vinte graus foram frequentes. Logo, não se podem generalizar resultados obtidos em experimentos com *depictions* para o que ocorre em cenas reais.

Os autores apontam ainda que a saliência visual da imagem está associada às fixações, de modo que áreas uniformes dos estímulos tendem a receber poucas fixações, enquanto áreas que se diferenciam das vizinhas são potencialmente informativas e são fixadas com mais frequência.

Griffin e Bock [5], buscando entender melhor os mecanismos cognitivos envolvidos na produção de sentenças, formularam um experimento de produção induzida com uso de monitoramento ocular. Quatro grupos de participantes participaram do experimento. O primeiro grupo descreveu eventos enquanto estes apareciam na tela, o segundo grupo produziu as sentenças depois que as imagens desapareciam da tela, o terceiro grupo recebeu a tarefa de observar a cena e encontrar o paciente da ação e o quarto grupo observou as imagens livremente. Os objetivos do experimento eram (i) verificar se os movimentos oculares seriam guiados por uma compreensão da cena ou pela saliência dos elementos; (ii) verificar se a apreensão da cena precederia a formulação sintática; (iii) verificar como a formulação e a execução estariam relacionadas. Os resultados revelaram que nos primeiros 1,330ms de exibição do estímulo não houve diferenças entre as fixações dos quatro grupos e os pacientes foram mais fixados que os agentes. Segundo as autoras, a similaridade dos movimentos oculares indica que as informações mais relevantes sobre a cena foram rapidamente extraídas, o que teria permitido aos indivíduos selecionarem o sujeito gramatical das sentenças com base na compreensão do evento e não na saliência da imagem.

Os pontos nos quais as fixações no agente e no paciente começaram a divergir foram comparados entre os grupos 1 e 3. Os resultados indicaram que não houve diferença significativa entre as duas condições: 288ms no grupo 1 e 316ms no grupo 3. A similaridade entre os pontos evidencia que a essência da ação representada na cena (*gist of the scene*) foi rapidamente extraída. O monitoramento dos movimentos oculares dos participantes do grupo 1 revelou direcionamento com base em um processo de formulação linguística, o que para Griffin e Bock é uma evidência de que o processo de produção da fala é iniciado com a conceptualização da mensagem e é procedido pela formulação incremental da sentença.

A medida do tempo revelou que os participantes passaram mais tempo fixando nos agentes antes de começarem a produzir o sujeito sintático da sentença do que durante a produção da frase. (646ms e 179 ms, respectivamente). No que tange às fixações no paciente, os participantes passaram mais tempo fixando esse elemento durante a produção da frase do que antes do início da produção do sujeito (812 e 244 ms). Houve uma forte ligação entre o número de fixações em uma área e o primeiro elemento mencionado, independentemente do tipo de estrutura linguística. Isso aponta para uma relação entre visão-mente-fala. Em conjunto, os resultados dos quatro grupos revelam que um processo holístico de conceptualização da cena acontece antes da formulação da fala.

Gleitman e colaboradores [6] replicaram o experimento de Griffin e Bock [5] fazendo uso de monitoramento ocular e da técnica de manipulação da atenção visual. Tal técnica consistiu na exibição de um asterisco situado no centro da tela por 500ms, seguido por um segundo painel, no qual havia um pequeno quadrado posicionado na região onde, posteriormente, seria exibido um dos elementos da cena. O quadrado permanecia na tela por 60 a 75ms e só então o estímulo visual era apresentado. Dois experimentos foram realizados. No primeiro, objetivou-se induzir frases que continham verbos de perspectiva e/ou sintagmas nominais combinados na posição de sujeito. No segundo experimento, foram analisados verbos de perspectiva, predicados simétricos e estruturas ativas e passivas. Porém, não foi utilizado o recurso de captação visual nas

imagens que eliciariam verbos de perspectiva. Em ambas as análises, houve efeito significativo do recurso de manipulação atencional na estrutura linguística produzida, indicando uma forte ligação entre a visão e processamento linguístico. Segundo os autores, a manipulação atencional favoreceu o acesso lexical a um dos termos. Verificou-se também similaridade entre os padrões dos movimentos oculares e padrões de fala, de modo que os elementos que foram fixados primeiramente também foram mencionados primeiro. Esses resultados apontam para um caráter mais incremental do processamento da fala: a apreensão da cena e construção sintática ocorreriam paralelamente.

O conceito de *gist of the scene* é fundamental para os trabalhos que investigam o curso incremental do processamento linguístico fazendo uso de imagens, como é o caso dos experimentos reportados. Henderson e Ferreira [4] compilaram resultados de vários estudos que tinham por objetivo explorar esse conceito. Um dos aspectos considerados é o tempo de reconhecimento do *gist*. Biederman e colaboradores, utilizando o paradigma de detecção de objeto (*object detection paradigm*), obtiveram evidências de que o *gist* é apreendido muito rapidamente. Na versão mais recorrente desse paradigma, o nome do objeto alvo é apresentado anteriormente à cena e posteriormente à breve apresentação da cena, um marcador de localização espacial é colocado em uma determinada localização e o participante deve identificar se o objeto estava ou não naquela localização. Os resultados de experimentos que utilizam esse paradigma apontam que as respostas dos participantes são predominantemente influenciadas pelo grau de consistência entre o objeto alvo e a cena apresentada, mesmo que o objeto nem esteja presente no estímulo observado. Para que esse tipo de relação de grau de consistência seja estabelecido é necessário que um número suficiente de informações sobre a cena tenha sido apreendido, o que corrobora a colocação de que o *gist of the scene* é rapidamente identificado.

Outro paradigma que vem sendo utilizado nos experimentos psicolinguísticos com o objetivo de investigar o rápido processamento do *gist of the scene* é o *Rapid serial visual presentation* (RSPV). Nesse paradigma, os participantes são apresentados a uma sequência de representações de cenas, na qual cada imagem permanece na tela por apenas 113ms. Em seguida, pergunta-se aos sujeitos se uma determinada cena estava presente na sequência (mostrando a cena alvo – *detection condition*) ou se havia na sequência uma cena com determinadas características (a cena alvo não é mostrada novamente, apenas descrita – *memory condition*). Os resultados mais significativos provêm da primeira condição e apontam que os participantes identificaram corretamente se as cenas alvo faziam ou não parte da sequência mesmo quando observaram cada cena por um tempo mínimo. Potter (1976) ampliou esses primeiros resultados e, segundo ele, o reconhecimento foi correto tanto quando o estímulo era descrito verbalmente como quando a imagem era apresentada. Isso sugere que o reconhecimento não depende de ter propriedades específicas da imagem alvo na mente. O que é mais interessante sobre esses resultados é o fato de que as cenas podem ser facilmente reconhecidas, porém não são lembradas pelos participantes, apontando que as cenas são rapidamente identificadas, mas depois se perdem na memória. Para que haja consolidação das imagens na memória é necessário um maior tempo de exibição das cenas.

Outros tipos de experimentos que evidenciam a rápida apreensão do *gist* apontam que, além da interpretação global da imagem, é a saliência visual de algumas regiões da cena que direciona os movimentos oculares iniciais. Estudo mais recentes revelam a rapidez de apreensão da essência da cena através de resultados direcionados para o fato de que fotografias apresentadas por 20ms podem ser corretamente categorizadas em supracategorias. Porém, mais uma vez a natureza do estímulo visual

pode ter afetado os resultados, uma vez que as fotografias utilizadas retratavam um único objeto e não um todo semanticamente organizado. Por essa caracterização ser comum a todos os tipos de estímulos geralmente utilizados em testes de Psicolinguística (*arrays*), é necessário investigar se nesses estudos a rápida apreensão do *gist of the scene* seria ou não possível. Várias hipóteses foram elaboradas para tentar explicar o fato de o *gist* ser apreendido mais rapidamente do que o tempo necessário para que os objetos de uma cena e suas relações sejam identificados. A primeira delas acredita que haja um dado objeto que é rapidamente identificado, a partir do qual a essência da cena é inferida. O ponto fraco dessa teoria está na possibilidade de o *gist* ser apreendido mesmo em cenas em que a imagem esteja embaçada e nas quais nenhum dos objetos possa ser diretamente identificado. A segunda hipótese sustenta a ideia de que existem alguns recursos no nível da cena (grandes formas e escalas volumétricas) que direcionam a extração do *gist*, sem que nenhum objeto específico precise ser identificado. Há ainda uma terceira hipótese que defende que a baixa frequência das informações espaciais presentes na periferia da cena determinaria o *gist*. A explicação mais recente para a rápida apreensão do *gist of the scene* sugere que os participantes fazem uso de representação holística da cena, denominada *spatial envelope*. Essa representação seria formada uma série de características da cena, tais como naturalidade, abertura, expansão, que não exigem uma análise hierárquica dos objetos. Tomadas em conjuntos, evidências de todas as hipóteses apresentadas sugerem que a categoria semântica de uma cena – *gist* – é identificada entre os 30 e 50 primeiros milésimos de segundo a partir da apresentação do estímulo visual.

Castelhano e Rayner [7] apontam que o *gist of the scene* é apreendido antes mesmo de os olhos começarem a se mover. Isso indica que a primeira fixação pode não ser decorrente da saliência visual e sim, de uma interpretação semântica do todo. Compreender esse conceito e seu tempo de apreensão é, portanto, essencial para investigar os estágios de produção de fala em tarefas de compreensão de cenas.

A literatura que aborda questões visão e compreensão de sentenças também explora a incrementalidade. Processos antecipatórios na compreensão incremental da sentença têm sido explorados a partir do chamado paradigma do mundo visual, em que o movimento ocular dos participantes é registrado por um equipamento de rastreamento ocular enquanto estes manipulam objetos a partir de um comando linguístico ou enquanto ouvem uma descrição sobre o que pode ocorrer com esses objetos. Altmann & Kamide [9] reportam que informação sobre complementos verbais é antecipada tão logo o verbo é apresentado. Por exemplo, numa sentença como *The boy will eat the cake*, os participantes dirigiam seu olhar ao bolo (em um *set* com outras figuras) já a partir do momento de apresentação oral do verbo. Informação relativa a tempo codificada no verbo também deflagrou processo de busca visual por elemento compatível. Altmann e Kamide [10] confrontaram sentenças como *The man has drunk all of the wine* e *The man will drink all of the beer* apresentadas juntamente com figuras de dois tipos: figura contendo um copo vazio (congruente com verbo no passado) e um copo cheio (congruente com verbo no futuro). Tanto na condição de futuro quanto na de passado, os participantes tornaram-se propensos a olhar mais para os alvos “congruentes” no início da expressão referencial. Esses resultados apontam para o papel de processos antecipatórios e para o caráter incremental do processo de compreensão da linguagem e colocam em discussão a natureza modular do processador.

Com base na literatura, é possível afirmar que há ainda muitas questões em aberto que precisam ser investigadas no que tange à interface linguagem-visão-ação. A presente pesquisa busca contribuir para os estudos nessa área.

Metodologia

A pesquisa em desenvolvimento apresenta caráter experimental. Para realização dos experimentos foram construídas imagens que pudessem eliciar a produção das estruturas linguísticas que se desejava investigar. Apresenta-se, a seguir, separadamente o método e os resultados de cada experimento.

Experimento 1

O primeiro experimento buscou verificar, a partir de um recurso de manipulação de atenção visual, se, ao se chamar a atenção para um dado elemento de uma cena, esse procedimento viria a afetar a escolha da estrutura linguística a ser empregada na descrição verbal da referida cena (por exemplo, se foco no personagem agente levaria à produção de frase na ativa).

Neste estudo, foi usada a técnica de produção induzida a partir da apresentação de imagens. A variável independente foi a natureza do elemento visual focalizado: agente/fonte vs. paciente/alvo, dando origem a duas condições: (c1) foco atencional no agente e (c2) foco atencional no paciente. A variável dependente foi o tipo de frase usada pelo participante para descrever a cena.

Método

Participantes:

Trinta alunos universitários com idade entre 18 e 27 anos participaram do experimento, cujos estímulos foram organizados em duas listas. Quinze participantes realizaram o experimento com a lista 1 de estímulos, na qual o foco atencional estava posicionado sobre o agente da ação, e quinze participantes o realizaram com a lista 2, na qual o recurso de manipulação visual situava-se no paciente da ação.

Materiais:

Foram utilizadas 36 imagens (4 de treino, 16 distratoras e 16 experimentais). Dos 16 estímulos experimentais, metade procurava eliciar a produção de sentenças em que fossem empregados verbos de perspectiva. Todas as imagens utilizadas no experimento passaram por uma avaliação prévia, realizada com quinze voluntários, a fim de verificar se os estímulos escolhidos elucidariam os tipos de estruturas que se objetivava testar.

Aparato:

Os estímulos foram dispostos sequencialmente em forma de slides, no programa PowerPoint, e exibidos aos participantes em um notebook. Utilizou-se a técnica de manipulação da atenção visual. O recurso foi similar ao utilizado nos estudos de Gleitman [4] – um pequeno quadrado apresentado previamente à imagem, posicionado onde seria projetado ou o agente ou o paciente. O recurso permanecia na tela por 500ms e, em seguida, a tela era substituída por outra com a cena a ser descrita. A posição – direita e esquerda – do foco atencional foi controlada (metade dos estímulos com foco à direita e metade à esquerda).

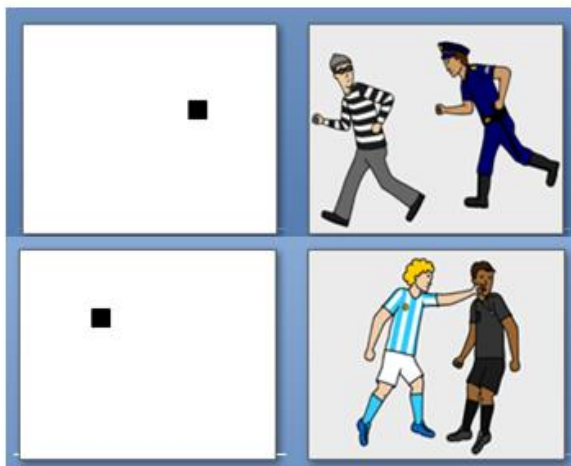


Figura 1: Exemplos dos estímulos utilizados na lista 1, com foco atencional no agente.

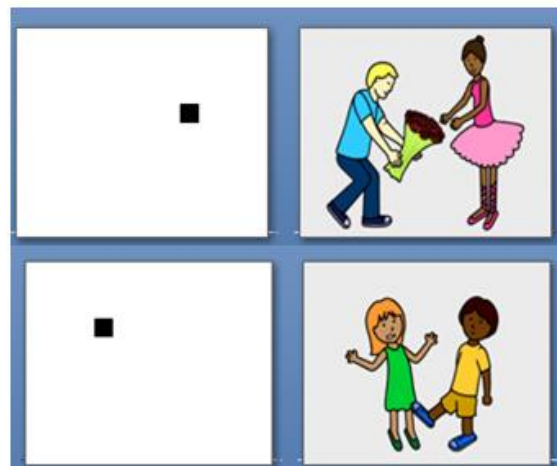


Figura 2: Exemplos de estímulos utilizados na lista 2, com foco atencional no paciente.

Como ilustrado nas imagens acima, na lista 1, o foco atencional posicionava-se sempre sobre o agente da ação e na lista 2, sempre sobre o paciente. Em alguns estímulos o sujeito estava à esquerda do vídeo (como na cena do jogador batendo no juiz e na do rapaz entregando flores à bailarina) e em outros, à direita (como na cena do policial perseguindo o ladrão e na do menino chutando a menina).

Procedimento:

Os participantes foram instruídos a observar as imagens e a produzir livremente sentenças que descrevessem a ação reproduzida em cada um dos desenhos. Ao fim de cada frase, eles apertavam um botão para que a próxima tela fosse exibida. As respostas foram gravadas com o auxílio de um gravador de voz do próprio notebook no qual os estímulos foram exibidos.

Resultados e discussão

No experimento 1, verificou-se que, independentemente da condição experimental, os participantes privilegiaram estruturas ativas. Os resultados foram submetidos à análise estatística por meio do teste *Mann Whitney unrelated*. Numa primeira análise, estruturas ativas e de verbos de perspectiva 1 foram contrastadas nas duas condições, tendo sido verificado um valor de p significativo ($p=0,0035$). Numa segunda análise, estruturas passivas e de verbos de perspectiva 2 foram contrastadas nas duas condições e o resultado também mostrou-se significativo ($p=0,0174$). Pode-se afirmar, portanto, que a natureza do elemento focalizado tem efeito sobre a produção de sentenças, visto que houve, na condição com foco no paciente, um menor número de estruturas ativas e um incremento na produção de passivas, em comparação à condição com foco no agente. Como ilustrado no gráfico a seguir:

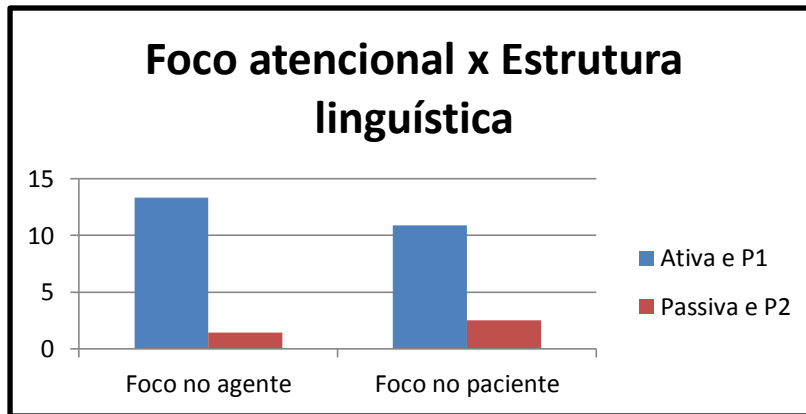


Gráfico 1: Resultados do experimento 1

Experimento 2

O segundo experimento buscou verificar a relação entre os pontos de fixação e a trajetória do olhar durante o mapeamento da cena visual e a estrutura linguística produzida. Para tanto, nesse estudo utilizou-se a técnica de rastreamento ocular (*eyetracker*).¹

Nesse experimento, aplicou-se a mesma tarefa do experimento anterior. Foi analisada a relação entre o tipo de frase usada pelo participante para descrever a cena e os seguintes fatores: a natureza do elemento fixado primeiro e do elemento fixado no *onset* da resposta verbal (agente/fonte vs. paciente/alvo), a latência da resposta verbal e a latência da primeira fixação.

Método

Participantes:

Até o momento, sete estudantes universitários com idades entre 23 e 47 anos participaram do estudo.

Materiais:

Das 36 imagens utilizadas no primeiro experimento, 10 foram selecionadas para o segundo experimento (2 de treino e 8 experimentais). Não foi utilizada nenhuma das imagens em que se buscava eliciar verbos de perspectiva.

Aparato:

O programa utilizado para a montagem do experimento foi o Tobii Studio3. As 10 imagens foram dispostas sequencialmente, sem o recurso da manipulação visual. Um equipamento *eyetracker* Tobii TX300 gravou os movimentos oculares dos participantes durante o experimento. As respostas dos sujeitos foram gravadas e analisadas no programa Sound Forge 8.

¹ Como um treinamento a respeito de questões de ordem metodológica relativas à utilização do equipamento, a aluna participou do workshop “Rastreamento ocular na pesquisa psicolinguística”, ministrado pelo professor Roberto Almeida (Universidade de Concordia, Canadá) e promovido pelo Programa de Pós-graduação Estudos da Linguagem da PUC-RIO e LAPAL/PUC-Rio, de 23-26 de outubro de 2012, como parte das atividades do I Ciclo de Debates em Linguagem.

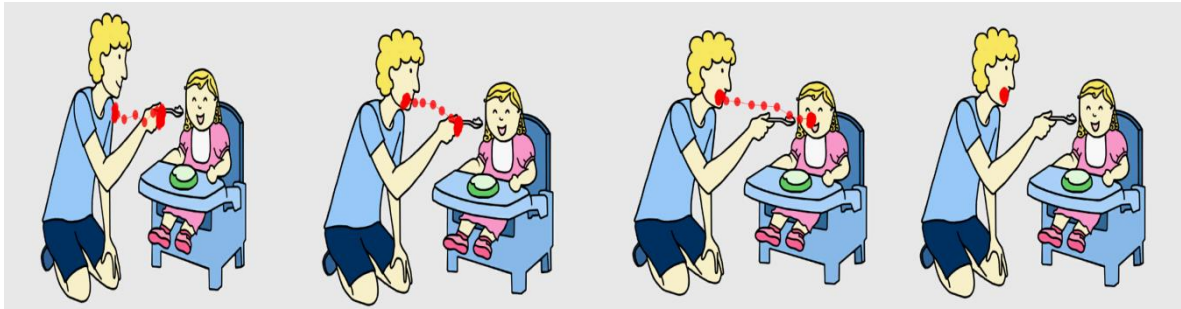


Figura 3: Imagens retiradas dos vídeos gerados pelo Tobii Studio como resultados do monitoramento ocular. Os círculos vermelhos representam as fixações. Nesse caso, a primeira fixação do participante foi no instrumento; essa foi seguida por fixações no agente e logo depois os olhos se moveram para o paciente. No onset da resposta verbal, o participante fixava o agente e produziu a frase “Um pai dando comida para a filha”.

Procedimento:

Os participantes foram instruídos a observar as imagens e a produzir o mais rapidamente possível sentenças que descrevessem a ação reproduzida em cada um dos desenhos. Ao fim de cada frase, eles apertavam um botão para que a próxima tela fosse exibida. Com objetivo de induzir que os participantes estivessem olhando para o centro da tela antes da exibição dos estímulos, era exibida, previamente à imagem, uma tela de transição, com três asteriscos ao centro.

Resultados e discussão

Resultados preliminares do experimento 2 vão de encontro aos resultados de Griffin e Bock [5], revelando preferência por estruturas ativas e pela fixação no paciente. O *onset* da resposta verbal ocorreu, em média, 1,7s depois da apresentação do estímulo visual, tempo bastante superior ao necessário à apreensão global da cena. Dessa forma, até agora, os resultados deste experimento parecem corroborar as conclusões de Griffin e Bock [5] de que os participantes terminariam o estágio da conceptualização da mensagem para só então dar início ao processo incremental de formulação sintática. Não foi observada relação entre o primeiro elemento fixado (agente/paciente) e a estrutura sintática selecionada (ativa/passiva). Parece haver, contudo, uma relação entre o elemento fixado no *onset* da resposta verbal e o tipo de frase produzida. Como ilustrado no gráfico a seguir:

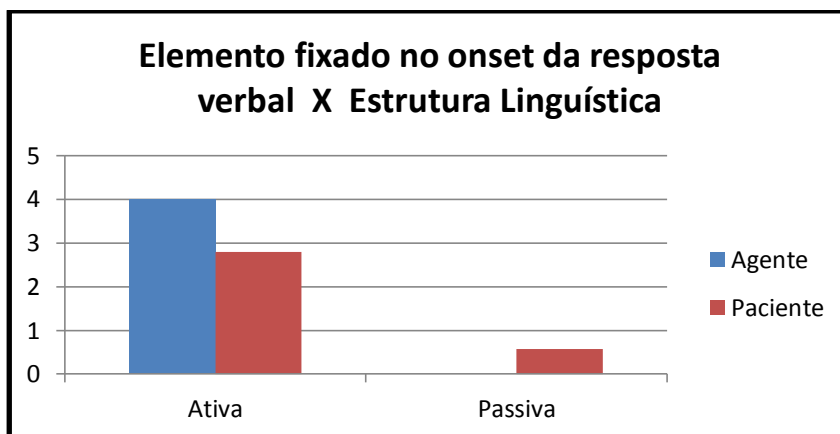


Gráfico 2; Resultados do experimento 2

O número de ocorrências de estruturas passivas foi baixo, contudo, em todas as vezes em que os participantes produziram esse tipo de estrutura linguística, o elemento fixado no *onset* da resposta verbal era paciente. No que tange às ocorrências de estruturas ativas, também parece haver relação com a natureza do elemento fixado no *onset* da resposta verbal. Das 52 ocorrências de sentenças ativas, 28 ocorreram quando o elemento fixado no *onset* da resposta verbal era agente, 20 quando era paciente. Em 4 ocorrências, a fixação ocorreu em elementos fora das áreas de interesse. O número de participantes é, não obstante, ainda pouco expressivo; faz-se necessário ampliar a amostra para verificar se a direção dos resultados se mantém e para se poder realizar tratamento estatístico dos dados.

Conclusões

O levantamento teórico sobre a interface linguagem-visão revelou que os cientistas que trabalham com visão pouco consideram as conclusões dos linguistas a respeito da busca visual e os especialistas em linguagem, por sua vez, não costumam preocupar-se com as propriedades visuais dos estímulos que utilizam em seus estudos.

Os resultados dos estudos experimentais desenvolvidos apontam que, no que tange à voz verbal, há forte preferência por estruturas ativas, que do ponto de vista do processamento são menos custosas. Em relação aos verbos de perspectiva, os resultados não são muito claros, sendo possível que fatores como frequência influenciem no acesso lexical a esses verbos e em seu conseqüente emprego na sentença.

Tomados em conjunto, os resultados dos dois experimentos conduzidos até o momento sugerem que, embora não se tenha observado uma correlação entre o elemento fixado nos instantes iniciais da inspeção da cena e a estrutura produzida, parece existir uma relação entre escolha estrutural e foco atencional (1º experimento)/elemento fixado contiguamente ao *onset* da sentença (2º experimento). Esses resultados, juntamente com o que já se tem conhecimento acerca da rapidez com que ocorre o processo de apreensão da essência de uma cena (*gist of the scene*), sugerem que, ao iniciar a formulação sintática, o participante já teve possibilidade de interpretar semanticamente a cena. Ele já pôde identificar qual é o predicador e os argumentos da proposição que expressa o conteúdo da cena, bem como já realizou o mapeamento do papel temático dos participantes (no caso, argumentos do predicador). A escolha propriamente dita do tipo de estrutura a ser produzida dependeria desse mapeamento inicial e também da definição do ponto de vista a partir do qual se deseja descrever a cena. Em relação à definição do ponto de vista, esta parece ocorrer posteriormente a identificação dos papéis temáticos e pode, como visto no experimento 1, sofrer interferência de fatores associados à atenção visual. O quanto outros fatores tanto visuais quanto linguísticos podem vir a influenciar a definição do ponto de vista e, conseqüentemente, da estrutura sintática mais adequada à sua expressão, é tópico que precisa ser ainda explorado a partir da manipulação de fatores como saliência visual, contexto prévio (com conseqüente definição de um tópico da mensagem), frequência lexical de um dado verbo de perspectiva em relação a seu correlato (exemplo: fugir vs. perseguir/ bater vs. apanhar), etc.

Nas próximas etapas da pesquisa, pretende-se aprofundar a investigação acerca do curso temporal na produção de sentenças iniciadas neste primeiro ano do projeto, fazendo uso, para isso, particularmente da técnica de rastreamento ocular. Questões relativas a processos antecipatórios na compreensão de sentenças por meio do paradigma do mundo visual também poderão vir a ser exploradas como contraparte aos estudos de produção.

Referências

- 1- RODRIGUES, E. dos S. **Processamento linguístico e incrementalidade: o que os olhos podem informar sobre o curso temporal da produção e compreensão de sentenças**. Programa Jovem Cientista do Nosso Estado (FAPERJ N° 17/2012).
- 2- FERREIRA, R; TANENHAUS, M. K. Introduction to the special issue on language–vision interactions. **Journal of Memory and Language**, 57, p. 455 – 459, 2007.
- 3-LEVELT, W. A Blueprint of the speaker. In C. Brown & P. Hagoort (Eds.). **The neurocognition of language**. Oxford Press, 1999.
- 4- HENDERSON, J.M FERREIRA, F. Scene Perception to Psycholinguists. In J. M. Henderson & F. Ferreira (Eds.) **The interface of language, vision, and action: Eye movements and the visual world**. New York: Psychology Press, 2004, p. 1–58.
- 5- GRIFFIN, Z. M.; BOCK, K. What the eyes say about speaking. **Psychological Science**, 11, p. 274–279, 2000.
- 6-GLEITMAN, L.R et al. On the give and take between event apprehension and utterance formulation. **Journal of Memory and language**, 57, p. 544–569, 2007.
- 7-CASTELHANO, M; RAYNER, K. Eye movements during reading, visual search and scene perception: an overview. In K.Rayner, D. Shem, X. Bai, & G. Yan (Eds). **Cognitive and Cultural Influences on Eye Movements**. Tianjin People's Press/Psychology Press, 2008.
- 9-ALTMANN, G.T.M ; KAMIDE, Y. Incremental interpretation at verbs: Restricting the domain of subsequent reference. **Cognition**, 73, 247–264, 1999.
- 10- ALTMANN, G.T.M ; KAMIDE, Y. The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. **Journal of Memory and Language**, 57, p.502-518, 2007.