

## **MODELOS PROBABILÍSTICOS E A RESOLUÇÃO DO PRONOME AMBÍGUO EM PORTUGUÊS BRASILEIRO**

PROBABILISTIC MODELS AND THE RESOLUTION OF AMBIGUOUS PRONOUNS  
IN BRAZILIAN PORTUGUESE

Mahayana Cristina Godoy  
Universidade Federal do Rio Grande do Norte  
mgodoy@cchla.ufrn.br

Matheus Araújo Mafra  
Universidade Federal do Rio Grande do Norte  
matmafra@gmail.com

### **Resumo**

Este artigo investiga como fatores gramaticais e informações pragmáticas atuam conjuntamente no estabelecimento da referência de um pronome pleno ambíguo em Português Brasileiro. Ao fazermos isso, nosso objetivo foi o de testar se o modelo baseado em expectativas de Kehler *et al.* (2008), que modela as relações entre produção e interpretação do pronome através do Teorema de Bayes, consegue explicar os dados dessa língua. Para tanto, realizamos um experimento de complementação de sentenças similar aos experimentos de Rohde (2008). Os resultados corroboram o modelo de Kehler *et al.* ao caracterizar interpretação e produção de pronomes como dois processos distintos que são influenciados, respectivamente, por pistas de natureza pragmática e gramatical.

**Palavras-chave:** pronome ambíguo, português brasileiro, relações de coerência, modelos probabilísticos, psicolinguística

This paper deals on how pragmatic information and grammatical factors interact while establishing the reference of ambiguous overt pronouns in Brazilian Portuguese. Our aim was to test whether Kehler *et al.* (2008) expectation-based model could explain our data by modulating the relations between pronoun production and pronoun interpretation through Bayes' Rule. In order to do that, we ran a sentence-completion task akin to the experiments seen in Rohde (2008). The results corroborate Kehler *et al.* model by distinguishing pronoun interpretation and pronoun production as two distinct processes that are influenced by pragmatic and grammatical cues respectively.

**Key-words:** ambiguous pronoun, Brazilian Portuguese, coherence relations, probability models, psycholinguistics

Recibido: 27/01/2018

Aceptado: 18/05/2018

## 1. Introdução

A compreensão da linguagem é dependente da nossa capacidade de relacionar sentenças entre si através do uso de elementos coesivos como anáforas e através da construção de relações coesivas entre as orações. Na sentença (1), abaixo, é nosso conhecimento do evento descrito que nos permite identificar o pronome "ele" como co-referente a "Caio". No entanto, no exato momento em que se depara com o pronome, o ouvinte/leitor ainda não tem elementos que permitam definir quem é seu antecedente. Se a sentença após o pronome fosse "... serviu o amigo com cuidado", possivelmente a leitura do pronome seria outra.

(1) Rodrigo serviu um bolo para Caio. Ele estava com fome.

Na psicolinguística, a resolução e produção de pronomes temporariamente ambíguos, como em (1), sempre foi um campo de estudo bastante explorado. O interesse da comunidade acadêmica no assunto se deve, em parte, pelo fato de que entender esse fenômeno lança luz sobre que pistas usamos para guiar o processamento *online* da linguagem. É essa a questão que procuramos investigar neste trabalho a partir de um estudo experimental de complementação de sentenças realizado em Português Brasileiro (doravante, PB) e inspirado no trabalho de Rohde (2008)<sup>1</sup>. Nossas hipóteses e análises constroem-se a partir do modelo proposto por Kehler *et al.* 2008 (doravante, Kehler *et al.*) para a resolução de pronomes ambíguos. A escolha desse modelo se deve ao fato de ele acomodar efeitos oriundos de informações pragmáticas e gramaticais e de pressupor que produção e interpretação pronominal são dois processos distintos, o que parece mais adequado para explicar a variedade dos dados que há hoje acerca da resolução de pronomes ambíguos.

Ao assumirmos o modelo de Kehler *et al.*, nosso objetivo principal foi o de testar se ele é capaz de explicar os dados do PB. A motivação para tal estudo se deve ao fato de o PB ter características específicas quanto ao uso de pronomes plenos de terceira pessoa do singular que permitem testar a confiabilidade do modelo de uma maneira que não seria possível em inglês, língua a partir da qual as hipóteses dos autores foram formuladas. Assim, devido à necessidade de estudos de processamento de linguagem buscarem evidências em outras línguas que não o inglês para poderem generalizar os seus achados (Norcliff *et al.* 2011), julgamos que os dados aqui documentados são de importância não só para a validação da proposta de Kehler *et al.*, mas para a compreensão do fenômeno da resolução pronominal em contexto de

---

<sup>1</sup> Esse trabalho inspira-se especialmente no Experimento I e Experimento VIII da autora.

ambiguidade. Nosso trabalho se inicia apresentando um breve resumo dos diferentes modelos que buscam compreender o processamento de pronomes ambíguos. Na seção 3, tratamos de como o modelo de Kehler *et al.* explica alguns dos dados sobre a interpretação de pronomes que parecem contraditórios, e levantamos algumas hipóteses sobre como seriam os dados do PB se realmente a interpretação de pronomes ambíguos puder ser modelada pelo Teorema de Bayes, conforme sugerem os autores. Por fim, apresentamos a replicação do trabalho de Rohde (2008) em PB e suas análises, discutindo como nossos dados colaboram para a validação do modelo testado.

## 2. Interpretação e produção de pronomes

Dentre os mais diversos modelos que tratam da resolução pronominal, é consensual a noção de que pronomes são geralmente utilizados para fazer referência à entidade mais saliente no discurso. Essa premissa está na base de muitos dos trabalhos que tratam da resolução de anáforas pronominais, ainda que o conceito de saliência seja definido de modo distinto através das noções de proeminência (Arnold 2001), acessibilidade (Ariel 1990), atenção (Gordon *et al.* 1993) ou centralidade (Grosz *et al.* 1995). Na esteira desse raciocínio, a resolução de pronomes ambíguos tem sido tratada como um processo de encontrar, entre os candidatos a referente, aquele que é mais saliente no contexto anterior. A distinção entre os modelos que buscam explicar a resolução de pronomes ambíguos tende a divergir justamente na definição de que tipo de pistas os falantes acessam para resolver a ambiguidade incrementalmente.

De um lado, há uma tradição que vê em informações providas pela estrutura gramatical a fonte principal de pistas que devem ser mobilizadas durante a interpretação do pronome. Uma das propostas que se baseiam nessa visão é a *Parallel Preference*, ou Preferência pelo Paralelismo, que advoga pela preferência em se resolver um pronome buscando seu antecedente na expressão referencial que ocupa a mesma posição gramatical na sentença anterior (Chambers e Smyth 1998).

Similarmente, algumas versões da *Centering Theory* também dão destaque a informações gramaticais na resolução pronominal. Essas propostas preveem que as entidades mencionadas no discurso constituem "centros" que guiam a resolução pronominal, e embora considerem que esse processo seja resultado de múltiplos fatores, a saliência desses centros é geralmente definida com base em informações gramaticais. Assim, geralmente sujeitos sintáticos são vistos como mais salientes que objetos, que, por sua vez, seriam mais salientes que referentes que ocupam outras posições sintáticas (Brennan *et al.* 1987).

Por outro lado, há modelos que tendem a caracterizar a resolução como produto de expectativas geradas a partir da representação discursiva do evento, atribuindo alguma ou nenhuma ênfase a pistas oriundas da estrutura gramatical.

A *Expectancy Hypothesis* (Arnold 2001), por exemplo, prevê que a resolução de pronomes é um processo de natureza probabilística em que atuam conjuntamente uma série de pistas que determinam a saliência de um referente. Nesse quadro, o pronome é interpretado a partir do antecedente em maior proeminência. A proeminência é então definida como a expectativa de que o referente será mencionado no contexto subsequente, e dentre a série de pistas que atuam para determiná-la estão pistas referentes à posição gramatical dos referentes, mas também informações da semântica do verbo. A ideia de expectativa como guia do processo de interpretação do pronome ambíguo é também encontrada na proposta de Hobbs (1979), com a diferença de que o autor nega que haja qualquer influência de pistas gramaticais atuando nesse processo. Em seu modelo, a interpretação do pronome é um subproduto de processos inferenciais mais gerais na construção de uma representação discursiva. Essas inferências estariam relacionadas ao estabelecimento de relações de coerência entre sentenças, e a resolução do pronome se daria pela definição de qual referente teria mais chances de ser mencionado no contexto seguinte dadas as restrições previstas por essas relações. Nesse cenário, a resolução do pronome dependeria das relações de coerência estabelecidas no discurso. Uma vez que se alterem essas relações, altera-se a leitura que se faz do pronome (cf. Wolf *et al.* 2004).

As divergências entre os modelos citados são várias no que tange às pistas utilizadas para a resolução de um pronome temporariamente ambíguo: alguns supervalorizam pistas gramaticais enquanto negligenciam informações contextuais, enquanto outros assumem uma posição inversa. Contudo, conforme notam Rohde e Keller (2014), esses trabalhos partem da premissa comum de que a produção e a interpretação de pronomes ambíguos são resultado de um mesmo processo. Nessa perspectiva, a produção pronominal é um reflexo da compreensão: falantes produzem um pronome para retomar um referente que está mais saliente, e ouvintes interpretam o pronome como pista para acessar o referente mais saliente no discurso. Embora intuitiva, essa visão de compreensão e produção de pronomes como resultado de um mesmo processo não é capaz de explicar alguns dados em que a produção e a interpretação de pronomes ambíguos em um mesmo contexto parecem depender de diferentes fatores.

Os dados obtidos por Rohde (2008) são um bom exemplo do limite de modelos que assumem uma simetria entre compreensão e interpretação de pronomes ambíguos. Em um experimento de complementação de sentenças que contou com textos como em (2a) e (2b), Rohde identificou que os participantes interpretavam o pronome a partir do referente no papel temático de alvo da oração anterior (Bob, em 2(a-b)) em taxas maiores para contextos perfectivos. Segundo a autora, tal comportamento deve-se ao fato de o aspecto perfectivo ressaltar a completude do evento, aumentando a probabilidade de uma menção ao alvo da oração seguinte. Contudo, o efeito do aspecto não foi tão robusto a ponto de determinar a preferência pela leitura de alvo em (2a).

Repetindo o achado de outros estudos (Stevenson *et al.* 1994), a autora identificou que os participantes não mostraram preferência por interpretar o pronome *He* como alvo ou fonte da oração anterior nos contextos perfectivos.

Nesses contextos, as taxas de interpretação do pronome como alvo ou fonte foram as mesmas, em torno de 50% para cada leitura.

Segundo a autora, o resultado poderia ser um reflexo de dois vieses: o primeiro, discursivo, indicando a fonte como referente mais saliente; o segundo, gramatical, associado ao uso de pronomes como expressões referencialmente utilizadas para fazer menção ao sujeito da oração anterior.

- (2) a. John handed a book to Bob. He \_\_\_\_\_  
b. John was handing a book to Bob. He \_\_\_\_\_  
c. John handed a book to Bob. \_\_\_\_\_  
d. John was handing a book to Bob. \_\_\_\_\_

Em um segundo experimento, Rohde (2008) documentou que a taxa de continuações que tinham o alvo como sujeito da oração (através do uso de um pronome ou qualquer outra expressão referencial) foi significativamente maior em (2c-d) do que em (2a-b), ainda que o efeito de aspecto se fizesse sentir nos dois grupos. Ao analisar os dados de produção de pronomes, Rohde identificou que a posição gramatical do antecedente, e não o aspecto verbal, foi fator decisivo no processo. Quando os participantes decidiam usar um pronome para continuar sua sentença em (2c-d), essa escolha era feita majoritariamente para referenciar o referente na posição de sujeito, independente do aspecto da oração, o que indicaria que a produção é influenciada por aspectos gramaticais, mas não por pistas oriundas da representação discursiva do evento.

Em resumo, esses dados indicam que a produção pronominal não é necessariamente influenciada pelas mesmas pistas contextuais que influenciam a interpretação dessas expressões. Essa incompatibilidade entre produção e compreensão de pronomes é observada em outros contextos (Stevenson *et al.* 1994; Fukumura e van Gompel 2010; Rohde e Horton, 2014; Kehler *et al.* 2008), e motiva o desenvolvimento de modelos que deem conta dessa disparidade.

### **3. A interação entre viés de produção do pronome e pistas discursivas no estabelecimento da interpretação de pronomes ambíguos: o modelo de Kehler *et al.* (2008)**

Indo ao encontro do documentado em Rohde (2008), Fukumura e van Gompel (2010) também encontraram evidências empíricas de que a produção e interpretação de expressões anafóricas são dois processos distintos. Em um experimento de compreensão e produção, os autores identificaram que informações semânticas oriundas do verbo e de conectivos entre sentenças influenciavam a interpretação, enquanto informação de cunho gramatical influenciava a escolha da forma da expressão utilizada (pronomes ou nomes).

Esse resultado ainda faz eco ao achado mais antigo de Stevenson *et al.* (1994), que compararam a continuação de sentenças como em (2a) e (2c) e concluíram haver influência de informações de naturezas diferentes na interpretação do pronome.

A proposta de Kehler *et al.* para lidar com essa aparente discrepância entre produção e interpretação caracteriza-se por ser um modelo baseado em expectativas.

Isso significa que os autores assumem que a interpretação pronominal é resultado de um processo de previsão sobre como o discurso deve seguir e que referente teria mais chances de ser mencionado. Essa proposta alinha-se com modelos mais gerais de processamento linguístico que ressaltam o papel da antecipação no processamento da linguagem para explicar fenômenos tão distintos quanto reconhecimento de fala e palavra (Salverda *et al.* 2014; Dikker e Pylkkänen 2013), processamento sintático (Fine *et al.* 2013) e produção/compreensão discursiva (Rohde e Horton 2014). Segundo alguns desses modelos, durante o processamento linguístico haveria a produção de previsões acerca de que informações (linguísticas, discursivas) seriam encontradas em um contexto subsequente, e o material linguístico atuaria atualizando essas expectativas de forma incremental (Kuperberg e Florian Jaeger 2015).

Dentro desse quadro mais geral, o modelo de Kehler *et al.* postula que expectativas são geradas por meio de processos inferenciais. Informações discursivas agiriam de modo a criar expectativas sobre como o discurso deve continuar em termos de relações de coerência e sobre que referentes seriam preferencialmente retomados caso uma dessas relações ocorresse (a exemplo de Hobbs (1979)). Dito de outro modo, haveria um cálculo da probabilidade de um referente ser mencionado -  $p(\text{referente})$  - independente da expressão referencial usada para retomá-lo. O processo de interpretação do pronome aconteceria na interação dessa expectativa com uma informação *bottom-up*, que codifica a probabilidade de, naquela língua, um interlocutor utilizar um pronome dado um referente específico (i.e.,  $p(\text{pronome}|\text{referente})^2$ ).

Dito de outra maneira, em uma frase como “*John handed the book to Bob. He...*” há expectativa de que o referente a ser mencionado na sequência seja *Bob*, que ocupa o papel temático de alvo. Ao encontrar o pronome “*he*”, essa expectativa é atualizada, uma vez que, em inglês, pronomes são produzidos com mais frequência para fazer referência ao sujeito da oração anterior.

Desse modo, o modelo de Kehler *et al.* prevê que a expectativa (gerada por informações discursivas) interage com um viés associado à produção de pronomes (influenciado por aspectos gramaticais) para determinar a interpretação de um pronome. A relação entre expectativa e viés de produção pode ser modelada, segundo os autores, nos termos do Teorema de Bayes<sup>3</sup> conforme a equação em (3):

(3)

$$p(\text{referente}|\text{pronome}) = \frac{p(\text{pronome}|\text{referente}) p(\text{referente})}{p(\text{pronome})}$$

<sup>2</sup> Em probabilidade condicional, o termo  $p(A|B)$  é lido como “a probabilidade de A dado B”, ou “a probabilidade de ocorrência de A dado que B tenha ocorrido”.

<sup>3</sup> O Teorema de Bayes é expresso da seguinte forma:  $p(A|B) = \frac{p(B|A)p(A)}{p(B)}$ . Na equação em (3),  $p(\text{pronome})$  refere-se à probabilidade de uso de um pronome e tem o papel de normalizar as probabilidades entre 0 e 1.

Na equação em (3),  $p(\text{referente}|\text{pronome})$  é a probabilidade de ocorrer um referente dado um pronome, ou seja, o resultado do próprio processo de interpretação pronominal.

Esse termo captura a probabilidade de que um pronome seja interpretado a partir de um referente específico.

O tratamento dessas relações através do Teorema de Bayes não explica apenas os dados de Rohde (2008), mas foi também capaz de modelar resultados semelhantes aos observados em experimentos que estudaram construções com verbos de causalidade implícita e voz passiva (Rohde e Horton 2014). Em todos esses casos, o termo  $p(\text{referente})$  sempre pareceu determinado por informações semântico-pragmáticas que influenciavam a interpretação do pronome. Ao mesmo tempo, a produção de pronomes indicada pelo termo  $p(\text{pronome}|\text{referente})$  nunca demonstrou ser influenciado por essas mesmas informações semântico-pragmáticas, mas seu comportamento sempre foi determinado por aspectos gramaticais.

Apesar de esses trabalhos terem sido todos feitos em língua inglesa, mais evidências obtidas através do estudo de processamento de segunda língua (Grüter *et al.* 2016) e de estudos em catalão (Mayol 2016), japonês (Ueno e Kehler 2016) e coerano (Kim *et al.* 2013) corroboram algumas características do modelo de Kehler *et al.* Kaiser (2013) aponta para a importância de se testar as previsões feitas a partir desse modelo em outras línguas, principalmente se considerarmos a existência de diferenças entre as línguas com relação à distribuição e uso de expressões pronominais, o que poderia causar grande variação na probabilidade com que os falantes dessas línguas usam um pronome. Como o sistema pronominal do PB difere do inglês por incluir pronomes plenos e nulos, essa língua oferece a oportunidade de testar o modelo de Kehler *et al.*, mais especificamente o papel e interação dos termos  $p(\text{referente})$  e  $p(\text{pronome}|\text{referente})$  no estabelecimento da referência de um pronome ambíguo. Para nosso trabalho, convém tratar de cada um desses termos separadamente, indicando como características do PB podem (ou não) mudar as probabilidades encontradas por Rohde (2008) de modo a corroborar ou invalidar as previsões feitas a partir do modelo de Kehler *et al.* (2008).

### **3.1. As relações de coerência no estabelecimento de $p(\text{referente})$**

O modelo de Kehler *et al.* prevê que as expectativas que definem  $p(\text{referente})$  são, em grande parte, determinadas pelas relações de coerência que se estabelecem entre as sentenças. Essa premissa apoia-se na proposta de Hobbs (1979) e conta com evidências empíricas que atestam que a resolução pronominal é em grande parte afetada por relações de coerência no momento do processamento linguístico (Wolf *et al.* 2004).

Nesse quadro, a resolução do pronome seria resultado (i) da probabilidade de ocorrência de uma relação de coerência específica no contexto subsequente e (ii) da estimativa de quão provável seria mencionar um referente específico na sequência (o que, por sua vez, é condicionada pela relação de coerência prevista). Os resultados do experimento de Rohde (2008) dão um exemplo claro dessa dinâmica.

No caso de sentenças com verbos perfectivos como em (2a), o foco colocado na completude da sentença levaria o leitor/ouvinte a esperar que o texto prossiga de forma a dar informações acerca do que aconteceu após o evento descrito.

Isso elevaria a possibilidade de ocorrer uma relação coesiva de *consecutividade*, que trata do que ocorreu após o evento narrado na primeira sentença. Por sua vez, a ocorrência de tal relação coesiva favoreceria as chances de o referente no papel temático de alvo ser mencionado, visto que o foco do fim do evento recai sobre esse referente.

Por outro lado, o aspecto imperfectivo levaria o leitor/ouvinte a esperar mais informações sobre o evento em si, visto que a opção foi por narrar um evento ainda em andamento. Isso levaria a um aumento da probabilidade de ocorrência de uma relação coesiva de *elaboração*, que provê mais informações sobre o próprio evento. Uma vez que o foco dessa relação de coerência recai sobre o evento, é esperado que sua ocorrência aumente a chance de se mencionar o agente dessa ação - no caso, o referente no papel temático de fonte.

Os dados de Rohde (2008) corroboram essas hipóteses e identificam a relação de coerência como a responsável por estabelecer a interpretação do pronome. A autora identificou que em relações de consecutividade, o pronome era preferencialmente interpretado como alvo da oração, independente do aspecto verbal da sentença. Para relações de elaboração, o pronome era lido preferencialmente como o referente à fonte da oração anterior, e mais uma vez não houve influência do aspecto. O que o aspecto verbal fez foi aumentar as taxas de relações de consecutividade para sentenças perfectivas comparativamente às imperfectivas, além de aumentar as relações de elaboração para as sentenças imperfectivas quando comparadas às perfectivas.

Os resultados da autora para as relações de coerência assentam-se em processos inferenciais mais gerais. Seus efeitos dependem de características do *input* linguístico que aumentem ou diminuam a probabilidade de ocorrência de determinadas relações (como o aspecto verbal, por exemplo), mas a base dessa inferência decorre da representação do evento e conhecimento semântico-pragmático. Por não se apoiarem em características próprias da gramática de uma língua específica, é possível que essas inferências tenham caráter mais universal entre as diferentes línguas, e que os achados do PB com relação ao papel do aspecto e relações de coerência na resolução do pronome espelhem os resultados reportados para o inglês. Essa é a hipótese que testaremos no experimento descrito na seção 4.

### **3.2. Diferenças tipológicas e o termo *p*(pronome|referente)**

No modelo de Kehler *et al.*, o termo *p*(pronome|referente) captura a influência do viés de produção, indicando a probabilidade de produzir um pronome para retomar um determinado referente. Dadas as diferenças de sistema pronominal entre as línguas naturais, há motivos para supor que há grande variação entre-línguas com relação ao valor atribuído a esse termo.

Para o inglês, por exemplo, há inúmeros estudos indicando uma preferência pela produção de pronomes como co-referente ao sujeito da oração anterior (Rohde 2008; Rohde e Keller. 2014). Contudo, isso possivelmente ocorre devido a características próprias do sistema gramatical dessa língua.

Em línguas românicas, advoga-se uma distribuição complementar no uso de pronomes plenos e nulos, postulando-se que pronomes nulos recuperariam preferencialmente o antecedente em posição de sujeito, enquanto pronomes plenos tenderiam a retomar o antecedente na posição gramatical de objeto (Carminati 2002). Com relação ao pronome nulo, diversos estudos de fato destacam uma preferência pela co-referência com o sujeito da oração anterior em PB (Almor *et al.* 2017; Teixeira *et al.* 2014), português europeu (Luegi 2012), italiano (Carminati 2002), espanhol e catalão (Alonso-Ovalle *et al.* 2002). Contudo, o uso e interpretação do pronome pleno é uma questão mais controversa.

Para o italiano, há evidências de que o pronome pleno é preferencialmente interpretado a partir do referente na posição de objeto da oração anterior (Carminati 2002); por outro lado, estudos do catalão e espanhol indicam que referentes na posição de sujeito e objeto são igualmente apontados como possíveis antecedentes de um pronome pleno (Alonso-Ovalle *et al.* 2002). Essa disparidade pode ser o resultado indireto dos diferentes métodos experimentais empregados pelos autores, mas indicam que a influência de aspectos gramaticais na interpretação do pronome pleno não parece tão homogênea entre diferentes línguas. De fato, estudos experimentais que fizeram uma comparação direta entre italiano e espanhol identificaram o mesmo comportamento para a interpretação de pronomes nulos entre as línguas, mas comportamentos distintos com relação à interpretação do pronome pleno (Filiaci *et al.* 2014; Filiaci 2011).

Em PB, essas divergências com relação à interpretação do pronome pleno persistem em trabalhos que usaram estímulos experimentais bastante semelhantes. Investigando sentenças como "Emília acenou para Teresa quando  $\emptyset$ /ela abriu a porta", o trabalho de Fonseca e Guerreiro (2012) identificou uma tendência por interpretar o pronome pleno como co-referente ao objeto da oração anterior. Contudo, em um experimento de rastreamento ocular que fez uso de frases muito semelhantes, Teixeira *et al.* (2014) não identificaram qualquer preferência na interpretação do pronome. Essa diferença reforça que a resolução do pronome pleno parece ser mais permeável a efeitos da tarefa experimental e/ou de características do estímulo não controladas pelos pesquisadores.

Tais resultados indicam que a probabilidade de uso de um pronome pleno para retomar o sujeito em PB (i.e.,  $p(\text{pronome pleno}|\text{sujeito})$ ) pode ser diferente do documentado em inglês, mas é preciso fazer algumas ressalvas entre os estudos sobre o PB aqui descritos e as predições feitas pelo modelo de Kehler *et al.* Em primeiro lugar, os resultados descritos focaram na interpretação dos pronomes, e não na sua produção. Portanto, segundo Kehler *et al.* eles não capturam exatamente  $p(\text{pronome pleno}|\text{sujeito})$ , mas sim  $p(\text{sujeito}|\text{pronome pleno})$ .

Além disso, esses estudos em PB - e boa parte dos estudos em outras línguas (Filiaci *et al.* 2014; Filiaci 2011; Carminati 2002) - tratam de contextos intrassentenciais, e é necessário ter cautela para generalizar tais resultados para os contextos inter-sentenciais como os investigados por Rohde (2008). Apesar dessas ressalvas, os resultados descritos podem indicar que o termo  $p(\text{pronome pleno}|\text{sujeito})$  para o PB tem um valor menor do que para o inglês, visto que não se documentou qualquer preferência pelo sujeito da oração anterior na interpretação do pronome pleno em PB. Se essa predição for verdadeira, em PB o termo  $p(\text{referente})$  teria um peso maior no estabelecimento da referência do pronome ambíguo do que o observado em inglês. Como resultado, seria possível que o efeito de aspecto fosse mais determinante para a interpretação do pronome em PB do que foi para o inglês.

Por fim, lembramos ainda que o modelo de Kehler *et al.* caracteriza a resolução de pronomes como parte de um processo de criação de expectativas acerca do contexto subsequente, e o termo  $p(\text{pronome}|\text{referente})$  teria papel importante em determinar essas expectativas. Assim, uma vez que a probabilidade de uso de um pronome para referir o sujeito da oração anterior seja muito alta, como é em inglês, a mera presença desse pronome faria como que o ouvinte/leitor atualizasse suas expectativas de modo a aumentar a probabilidade de que o sujeito seja mencionado na sequência. Como a referência do pronome é determinada pelas relações de coerência entre as sentenças, em inglês a presença do pronome aumentaria as taxas de relações de elaboração em detrimento das relações de consecutividade, o que explicaria as diferenças de menção à fonte em comparação a contextos em que um pronome não ocorre. De fato, esse é o cenário observado no trabalho de Rohde (2008), o que corrobora que a presença do pronome é fator que introduz um viés responsável por recalculas as expectativas sobre como o discurso deve se desenvolver.

De modo geral, o que essas predições e os estudos discutidos aqui mostram é que o modo como a posição sintática do antecedente influencia a interpretação do pronome pleno parece ter grande variabilidade entre línguas. Portanto, tomando como premissa o modelo de Kehler *et al.*, é possível imaginar que as línguas difiram com relação ao viés introduzido pelo termo  $p(\text{pronome}|\text{referente})$ . Isso teria um impacto nas relações de coerência esperadas (e produzidas) pelo ouvinte/leitor, e por consequência seria uma possível causa das diferenças observadas na interpretação do pronome pleno entre diferentes línguas. O experimento que descrevemos a seguir tem por objetivo testar o modelo dos autores, identificando se os padrões observados em PB também podem ser explicados por ele.

#### **4. Experimento: objetivos e hipóteses**

Na última seção, levantamos hipóteses acerca de como seria a interpretação de pronomes plenos ambíguos em PB se  $p(\text{referente})$  e  $p(\text{pronome}|\text{referente})$  realmente atuam de acordo com o modelo de Kehler *et al.* Isso nos permite fazer algumas previsões e testar o modelo dos autores a partir de dados de uma língua tipologicamente distinta do inglês.

Com essa finalidade, conduzimos um experimento de completção de sentenças similar ao de Rohde (2008). A partir de sentenças como as apresentadas na Tabela 1, em que controlamos aspecto e tipo de estímulo, analisamos como os participantes continuavam as sentenças experimentais.

Isso nos permitiu investigar (i) como os pronomes ambíguos nas sentenças com estímulo pronominal são preferencialmente interpretados, e se essa interpretação se relaciona com o tipo de relação de coerência estabelecida; (ii) como aspecto e tipo de estímulo influenciam a escolha do sujeito da oração subsequente; (iii) que fatores influenciam a produção do pronome pleno em PB.

Rodrigo serviu a torta para Caio. Ele _____	Aspecto Perfectivo Estímulo Pronominal
Rodrigo estava servindo a torta para Caio. Ele _____	Aspecto Imperfectivo Estímulo Pronominal
Rodrigo serviu a torta para Caio. _____	Aspecto Perfectivo Sem estímulo
Rodrigo estava servindo a torta para Caio. _____	Aspecto Imperfectivo Sem estímulo

**Tabela 1.** Exemplo de sentenças experimentais

A primeira questão que pretendemos investigar diz respeito à influência de aspectos semântico-pragmáticos na resolução do pronome em PB. Assim, perguntamo-nos se a interpretação de pronomes plenos ambíguos nessa língua está sujeita à influência dos mesmos fatores pragmáticos identificados por Rohde (2008) para o inglês. Em caso afirmativo, devemos ver os mesmos efeitos de aspecto e relações de coerência que Rohde encontrou para o inglês. Aspectos perfectivos levariam a maior número de relações de consecutividade, que por sua vez aumentariam a taxa de leitura do pronome como co-referente ao alvo da oração anterior. Para sentenças imperfectivas, o maior número de relações de elaboração elevaria as taxas de co-referência entre pronome e fonte da oração anterior. Assim, deveríamos observar o efeito de aspecto verbal na mudança de distribuição das relações de coerência. Se o termo  $p(\text{pronome}|\text{fonte})$  em PB realmente for menor que em inglês, ainda devemos observar taxas de retomada do alvo da sentença experimental maiores do que os 50% documentados em inglês.

Para investigar possíveis diferenças entre inglês e PB com relação ao termo  $p(\text{pronome}|\text{referente})$ , investigamos se a produção do pronome pleno em PB também é influenciada pelo papel gramatical do antecedente. Conforme vimos, alguns estudos indicam que a interpretação do pronome pleno em PB não é afetada pela posição gramatical do antecedente (Teixeira *et al.* 2014), enquanto outros indicam que ele tende a retomar o antecedente na posição de objeto (Fonseca e Guerreiro 2012). Contudo, esses estudos não trazem dados sobre a produção de pronomes em um contexto experimental controlado.

Nosso trabalho poderá investigar essa questão ao analisar as continuações iniciadas por pronomes plenos (e nulos) em contextos sem estímulo. É importante ressaltar que nosso desenho experimental não permite distinguir papel temático e posição gramatical, mas supomos que qualquer viés de produção associado ao pronome pleno se deva a posição gramatical devido ao fato de ser um efeito fartamente documentado em outras línguas.

Os dados do PB nos levam a crer que não haverá preferência por produzir um pronome associado à posição, mas não é possível determinar, com base nos estudos anteriores, se haveria preferência pela produção de pronomes plenos para retomar o referente na posição de objeto da sentença anterior.

Por fim, perguntamo-nos se há evidência de que há vieses diferentes atuando na produção e na interpretação de pronomes, como defendem Kehler *et al.*, e se a interpretação do pronome ambíguo pode ser caracterizada como um processo de atualização de expectativas. Se a produção e compreensão do pronome são dois processos distintos, é possível que pistas que influenciam a interpretação do pronome pleno não influenciem sua produção.

Assim, os efeitos de aspecto que prevemos atuando na interpretação pronominal não devem determinar a produção pronominal. Além disso, os dados de Rohde (2008) documentam que a presença de um pronome, comparado a contextos sem estímulo pronominal, aumenta as taxas de continuações que fazem referência à fonte da oração anterior. Mais que isso, esse aumento seria resultado da mudança na distribuição das relações de coerência, o que influenciaria as taxas distintas de menção a fonte ou alvo no contexto subsequente.

Segundo a autora, isso seria evidência de que a resolução de pronomes é parte de um processo de criação de expectativas acerca do contexto seguinte, e que a mera presença dessa expressão referencial contribuiria para a atualização dessas expectativas. Se isso é verdade, qualquer diferença na preferência por fonte ou alvo motivada pelo tipo de estímulo em nosso experimento deverá ser explicada por diferentes distribuições nas relações de coerência estabelecidas entre as sentenças.

#### **4.1. Materiais**

Foram construídos 18 textos com evento de transferência de posse (TdP) que se desdobravam nas condições previstas na Tabela 1. Elas consistiam de uma sentença contextual seguida por uma sentença incompleta, no caso de itens com estímulo pronominal, ou apenas uma sentença contextual, no caso dos itens sem estímulo. As sentenças contextuais retratavam um evento de TdP no qual o aspecto verbal (perfectivo ou imperfectivo) foi manipulado. Os participantes viam apenas uma das quatro condições previstas para cada sentença. Nenhum verbo de TdP foi apresentado mais de uma vez ao mesmo participante.

Os referentes de fonte e alvo eram nomes próprios do mesmo gênero, criando a ambiguidade que servia de estímulo para que os participantes completassem os textos na condição de estímulo pronominal. O gênero dos pronomes nas sentenças com estímulo pronominal foi balanceado: metade apresentava o pronome feminino "ela" e a outra metade o pronome masculino "ele". A fim de evitar que a ambiguidade fosse entendida como referindo-se ao tema da sentença anterior (como "torta", na Tabela 1), os objetos inanimados escolhidos para esse papel tinham gênero gramatical diferente dos referentes em posição de fonte e alvo.

Além dos 18 itens experimentais, foram elaboradas 29 sentenças distratoras.

Elas consistiam de sentenças contextuais com verbos sem transferência de posse, e foram apresentadas com ou sem estímulo posterior, da mesma maneira que os itens experimentais. Entretanto, os estímulos apresentados para continuação das sentenças eram advérbios, nomes próprios e pronomes não-ambíguos. A partir desses itens, foram criadas quatro listas contendo as 29 sentenças distratoras e 18 itens experimentais distribuídos em um quadrado latino.

## 4.2. Participantes e tarefa

Participaram do experimento 128 falantes nativos de PB, 90 do sexo feminino, com idade média de 25 anos (idade mínima = 18 anos, idade máxima = 20 anos, desvio padrão = 4,5 anos). Todos os participantes tinham ensino superior completo ou incompleto e deram consentimento formal para participarem voluntariamente da pesquisa.

Os participantes foram instruídos a escrever continuações para os textos usando a primeira ideia que lhes ocorresse, evitando humor. Os estímulos foram apresentados a partir de um formulário *online* que os participantes puderam acessar de suas casas. Depois de preencherem um formulário de identificação com dados pessoais de idade, localização, língua materna e escolaridade, os participantes completaram três passagens como treinamento para reconhecimento da tarefa. Todos os participantes completaram a tarefa em menos de uma hora.

## 4.3. Codificação dos dados

Foram excluídos os dados de dois participantes por desrespeitarem as instruções. Do total de 2.304 continuações restantes, metade consistia de continuações de sentenças iniciadas por um pronome e a outra metade seguia sentenças sem estímulo. Os dados foram analisados pelos dois autores do artigo, que anotaram (a) a referência do sujeito da oração criada pelos participantes (fonte, alvo, ambígua entre fonte ou alvo ou outro referente); (b) o tipo de expressão referencial usada na posição de sujeito para as sentenças de estímulo livre (pronome pleno, pronome nulo, nome próprio ou outros); (c) as relações de coerência estabelecidas entre as orações, usando os critérios expostos em Rohde (2008).

Notamos aqui que a classificação das relações de coerência obedeceu aos critérios elaborados por Rohde (2008) para 6 relações distintas: consecutividade, elaboração, explicação, violação de expectativa, resultado e paralelismo. No entanto, para nossas análises de relação de coerência foram consideradas apenas as relações de consecutividade e elaboração. Isso foi feito por serem essas as duas relações de coerência mais comuns em nosso conjunto de dados e no de outros estudos (Rohde 2008; Grüter *et al.* 2016; Ueno e Kehler 2016) e por serem as estudadas por Rohde (2008, experimento VIII) para investigar o viés introduzido por um pronome na distribuição das relações de coerência. Seguindo as definições da autora,

"[relações de] elaboração foram definidas como aquelas que dão mais detalhes sobre o evento ou estado descrito na sentença. Tais detalhes incluíam informações adicionais sobre tempo, local, instrumento ou propriedades benefactivas do evento. (...) Relações de consecutividade foram definidas como aquelas em que era possível estabelecer uma relação temporal entre as duas sentenças, de modo que o evento da primeira sentença precedia o evento da segunda sentença." (Rohde 2008: 199-200)

Em nossos dados, alguns exemplos de relação de elaboração foram "Juliana estava levando o café para Juliana. No meio do caminho lembrou que esqueceu do açúcar" e "Rodrigo serviu a torta para Caio. Ele tinha feito uma receita especial de sua avó". Para as relações de consecutividade, tivemos textos como "João estava trazendo a camisa para João. Depois entregou a camisa para ele" e "Carol enviou o e-mail para Cláudia. Ela respondeu logo em seguida".

Do total de 2.304 dados experimentais, excluíram-se 218 casos (9.4%) pelo fato de ao menos um dos pesquisadores ter considerado ambígua a leitura do sujeito da continuação. Também foram excluídos 5 itens (0.02%) por serem ambíguos quanto ao tipo de relação de coerência estabelecida entre as sentenças, 72 observações (3.1%) que se referiam conjuntamente aos referentes da primeira oração através de pronomes plurais ou expressões coordenadas (e.g., "eles", "ambos", "os amigos"), 188 continuações (8.1%) que foram escritas como orações subordinadas (e.g., "que gostou da torta"), 9 casos de continuações *nonsense* (0.03%), 5 observações que partiam de leituras alternativas da sentença (0.02%) e 296 continuações (12.8%) em que o sujeito da continuação não era o referente no papel de fonte ou alvo da sentença experimental. Também excluímos 17 continuações (0.07%) que usaram descrições definidas, pronomes possessivos ou demonstrativos para inserir o sujeito da oração. Desse modo, as análises foram conduzidas em um conjunto de 1492 observações: 538 observações para as sentenças sem estímulo e 954 observações para as sentenças de estímulo pronominal. Em todos esses casos, o sujeito da oração produzida pelos participantes fazia referência clara ao alvo ou fonte da sentença contextual, e no caso das sentenças livres essa referência era feita através do uso de um pronome nulo, pronome pleno ou repetição do nome próprio, as três formas referenciais mais comuns em nosso conjunto de dados.

## 5. Análise dos resultados

Todas as análises foram feitas a partir modelos de regressão logística mistos com o pacote lme4 do *software* estatístico R, e a significância dos termos previstos nos modelos foram aferidas por meio de comparação de modelos aninhados (Baayen *et al.* 2008). Os p-valores reportados foram corrigidos para múltiplas comparações usando o método de *false-discovery rate* (Benjamini e Hochberg 1995). Como recomendado, assumimos a estrutura máxima de efeitos aleatórios, incluindo interceptos aleatórios para sujeitos e

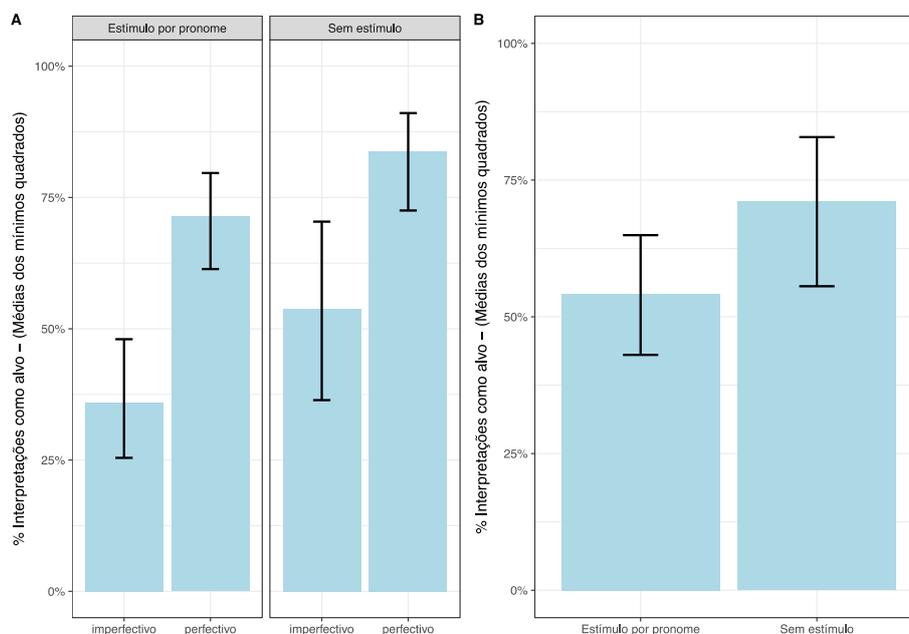
itens em todas as análises, além de coeficientes angulares aleatórios para aspecto e estímulo sempre que permitido pelo conjunto de dados (Barr *et al.* 2013).

### 5.1. Análise de aspecto e estímulo

Nossa primeira análise considerou como aspecto e tipo de estímulo, bem como sua interação, atuou na escolha do referente (fonte ou alvo) que figuravam como sujeito da continuação. As variáveis preditoras de aspecto e de estímulo foram centradas usando *deviation coding* (0.5, -0.5) para testarmos seus efeitos principais (Barr 2013). A interação entre aspecto e estímulo não adicionou contribuição significativa para o modelo ( $\chi^2(1) = 4.1244$ ,  $p = 0.056$ ), mas tanto estímulo ( $\chi^2(7) = 36.643$ ,  $p < 0.0001$ ) quanto aspecto ( $\chi^2(7) = 119.03$ ,  $p < 0.0001$ ) contribuíram significativamente. Os coeficientes obtidos a partir do melhor modelo ajustado são mostrados na Tabela 2, que indica maior ocorrência de menção ao alvo em aspectos perfectivos e em sentenças sem estímulo pronominal. A Figura 1 mostra as probabilidades estimadas pelo modelo.

	$\beta$	SE	Z	p-valor
(Intercept)	0.900	0.269	1.982	0.058
estímulo	0.732	0.227	2.220	<b>0.002</b>
aspecto	1.491	0.162	9.199	<b>&lt;0.001</b>

**Tabela 2.** Efeito de estímulo e aspecto na escolha referencial do sujeito da continuação para o melhor modelo ajustado (interpretação  $\sim$  aspecto + estímulo + (1+ estímulo + aspecto|participante) + (1+ estímulo + aspecto|item))



**Figura 1.** Probabilidade de continuação da sentença mencionando o alvo por aspecto e estímulo (as barras indicam intervalo de confiança de 95%)

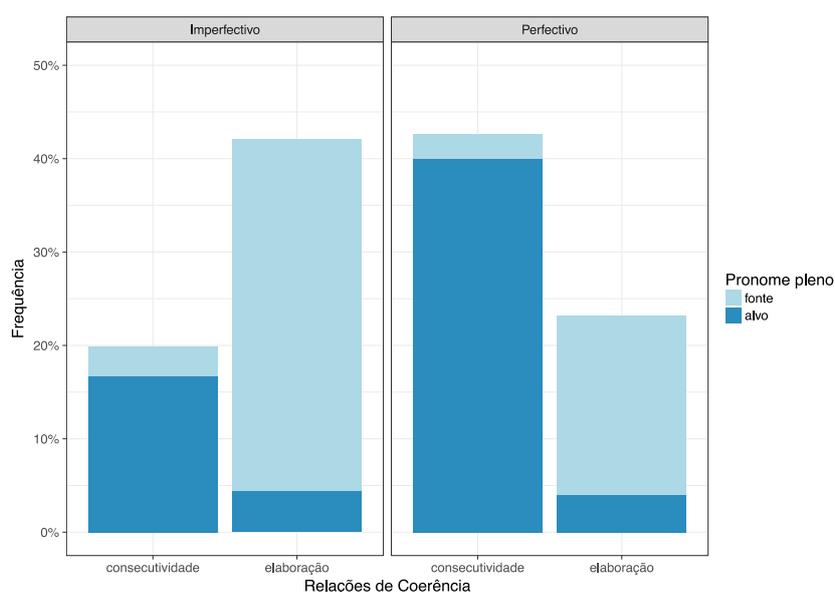
## 5.2. Análise dos dados de sentenças com estímulo de pronome pleno

Em nossa segunda análise, consideramos apenas o conjunto de dados em que havia estímulo pronominal ( $n = 954$ ). Em primeiro lugar, analisamos se as relações de coerência determinam a leitura do pronome em PB independentemente do aspecto verbal.

Para tanto, nossas análises restringiram-se às relações de consecutividade ( $n = 304$ ) e elaboração ( $n = 306$ ). Os dados provenientes de cada relação de coerência foram considerados separadamente.

A variável resposta dos modelos desenhados para cada conjunto foi a interpretação dada ao pronome, fonte ou alvo, e a variável preditora foi o aspecto verbal. Identificamos que o aspecto verbal não contribuiu significativamente para a interpretação dos pronomes em relações de consecutividade ( $\chi^2(1) = 0.3308$ ,  $p = 0.664$ ) e elaboração ( $\chi^2(1) = 0.1058$ ,  $p = 0.784$ ). Esses dados indicam que a preferência por fonte ou alvo, consideradas em cada uma das relações de coerência, não foi afetada pelo aspecto verbal. A Figura 2 mostra a frequência observada de leitura como fonte ou alvo do pronome para as relações de coerência em cada um dos aspectos verbais. Para avaliar se havia preferências significativas na interpretação do pronome como fonte e alvo em cada uma das relações de coerência, os dados dos contextos perfectivos e imperfectivos foram analisados separadamente.

Dados das relações de consecutividade mostram preferência pela interpretação do pronome como fonte acima de uma média hipotética de 0.5 em contextos perfectivos ( $n=215$ ,  $\beta = 10.07$ ,  $z = 4.534$ ,  $p < 0.0001$ ) e imperfectivos ( $n=89$ ,  $\beta = 1.7303$ ,  $z = 4.843$ ,  $p < 0.0001$ ). Em sentenças com relações de elaboração, os pronomes foram interpretados como se referindo preferencialmente à fonte independente do aspecto verbal (perfectivo:  $n=117$ ,  $\beta = -11.240$ ,  $z = -2.564$ ,  $p = 0.01$ ; imperfectivo:  $n=189$ ,  $\beta = -28.115$ ,  $z = -4.42$ ,  $p < 0.0001$ ).



**Figura 2.** Taxas de ocorrência de relações de consecutividade e elaboração por aspecto verbal e viés de interpretação do pronome pleno ambíguo

Por fim, investigamos se o aspecto verbal influenciou as taxas de ocorrência dessas relações de coerência. Para cada uma delas foi ajustado um modelo de regressão logística a partir do conjunto de todos os dados de estímulo pronominal.

A variável resposta foi a ocorrência ou não da relação de coerência avaliada e variável preditora foi o aspecto verbal. Através da comparação de modelos aninhados, identificamos que o aspecto contribui para o modelo das relações de consecutividade ( $\chi^2(1) = 44.253$ ,  $p < 0.0001$ ) e elaboração ( $\chi^2(1) = 35.173$ ,  $p < 0.0001$ ).

Conforme indicam os coeficientes obtidos pelo modelo, o aspecto perfectivo aumenta a probabilidade de ocorrência de relações de consecutividade em relação ao imperfectivo ( $\beta = 1.327$ ), e diminui a ocorrência de relações de elaboração ( $\beta = -1.035$ ).

### **5.3. Análise de taxas de expressão referencial nos dados de estímulo livre**

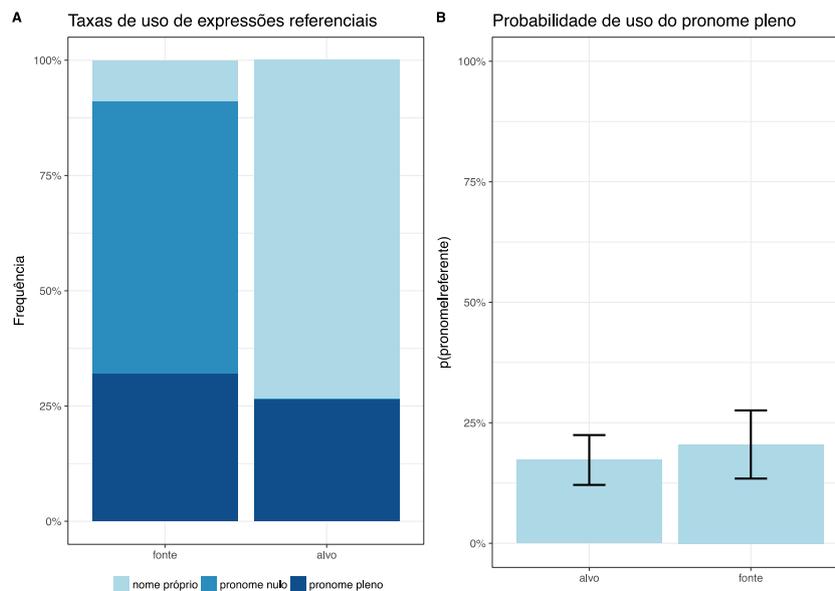
Nosso terceiro conjunto de análises focou nos dados dos textos sem estímulo, e investigou se as taxas de uso das três expressões referenciais analisadas - pronome pleno, pronome nulo ou nome próprio - indicam preferência do uso dessas expressões para se referir à fonte ou ao alvo da oração anterior.

Para a análise de cada expressão referencial foi considerada como variável dependente um fator de dois níveis indicando se a expressão em análise foi ou não foi usada. Como variável preditora em todos os modelos foi usada a referência estabelecida pela expressão referencial: fonte ou alvo.

A referência não contribuiu significativamente para o modelo do pronome pleno ( $\chi^2(1) = 0.1546$ ,  $p = 0.725$ ), mas contribui para o modelo do pronome nulo ( $\chi^2(1) = 229.71$ ,  $p < 0.0001$ ) e para o modelo do nome próprio ( $\chi^2(1) = 65.139$ ,  $p < 0.0001$ ).

Os coeficientes indicam que há mais produção de nomes próprios para se referir ao alvo em comparação à fonte ( $\beta = 8.100$ ), e que há mais produção de pronome nulo para se referir à fonte do que ao alvo ( $\beta = 14.711$ ).

Essas análises podem ser vistas na Fig 3, que mostra as taxas observadas para cada expressão referencial (cf. Fig 3A) e a probabilidade de uso de um pronome pleno para referenciar fonte ou alvo segundo o modelo ajustado (Fig 3B).



**Figura 3.** Relações entre forma referencial usada e referência pretendida: taxas observadas (Fig 3A) e probabilidades calculadas para uso do pronome pleno (Fig 3B) (as barras indicam intervalos de confiança de 95%)

Avaliamos ainda se o aspecto gramatical - fator que influenciou a interpretação do pronome pleno - poderia ser considerada variável preditora na produção de pronomes plenos. A produção de pronomes plenos foi codificada como uma variável resposta de dois níveis e a variável preditora foi o aspecto verbal. Uma comparação de modelos aninhados identificou que a inclusão de aspecto não apresenta contribuições significativas ( $\chi^2(1) = 0.0429$ ,  $p=0.836$ ).

#### 5.4. Influência da ocorrência do pronome na mudança das relações de coerência

Por fim, analisamos se o tipo de estímulo mudou a distribuição das relações de coerência de consecutividade e elaboração. Foi ajustado um modelo para cada relação de coerência, com a ocorrência ou não da relação como variável dependente e o tipo de estímulo como variável preditora. O tipo de estímulo foi significativo para o modelo no caso das relações consecutivas ( $\chi^2(1) = 4.854$ ,  $p = 0.03$ ), que ocorreram em maior proporção com os estímulos sem pronome ( $\beta = 0.326$ ) e para as relações de elaboração ( $\chi^2(1) = 6.281$ ,  $p = 0.01$ ), que ocorreram em menor proporção com os estímulos sem pronome ( $\beta = -0.3903$ ).

## 6. Discussão dos resultados

Nossa primeira pergunta de pesquisa indaga se, em PB, a interpretação de pronomes plenos ambíguos estaria sujeita à influência dos mesmos fatores pragmáticos identificados por Rohde (2008). Com relação aos efeitos do aspecto verbal, nossos dados indicam que o aspecto perfectivo leva a maior número de leituras do pronome como co-referente ao alvo da sentença

anterior em comparação com o imperfectivo. Esse resultado é similar aos achados de estudos feitos em inglês (Rohde 2008; Stevenson *et al.* 1994).

A diferença principal entre PB e inglês está no fato de que, nas sentenças de aspecto perfectivo, em PB há preferência por interpretar o pronome como co-referente ao alvo da oração anterior. Interpretamos esse dado como indício de que a probabilidade de usar um pronome para se referir à fonte da oração anterior (i.e., o termo  $p(\text{pronome pleno}|\text{fonte})$ ) tem valores distintos em PB e em inglês. Em PB, conforme argumentaremos adiante, não há influência da posição gramatical do antecedente no uso de um pronome pleno, enquanto em inglês esses pronomes são mais usados para fazer referência ao sujeito da oração anterior (Rohde, 2008). Por isso, é possível imaginar que as diferenças entre as línguas nas taxas de interpretação no aspecto perfectivo derivem do fato de que  $p(\text{pronome pleno}|\text{fonte})_{PB} < p(\text{pronome pleno}|\text{fonte})_{\text{inglês}}$ .

Adicionalmente, nossas análises sobre as relações de coerência indicam que a maior proporção de leituras de alvo para o pronome pleno em contextos perfectivos não é um efeito direto do aspecto verbal, mas um efeito indireto do tipo de relação de coerência que sentenças no perfectivo costumam selecionar.

O aspecto perfectivo tende a salientar o fim do evento, o que levou os participantes a escolherem relações de coerência como a de consecutividade, que indicam o que ocorreu na sequência do evento. Na ocorrência dessa relação de coerência, a leitura do pronome foi, preferencialmente a de alvo. Por outro lado, o aspecto imperfectivo salienta o desenrolar da ação, aumentando a probabilidade de ocorrência de relações que focam o evento, como a elaboração. Em um contexto de transferência de posse, enfatizar o evento em andamento aumentou a probabilidade de mencionar a fonte no contexto subsequente. Portanto, são as relações de coerência que determinam a interpretação do pronome pleno tanto em PB quanto em inglês, cabendo ao aspecto o papel de criar expectativas acerca de qual relação de coerência deve seguir no contexto subsequente.

A segunda pergunta que buscamos responder diz respeito ao uso de pronomes plenos para fazer referência à fonte - e, nos contextos de TdP, ao sujeito gramatical - da oração anterior. Alguns dos nossos dados vão ao encontro de estudos anteriores ao identificar o uso de pronomes nulos para referir-se preferencialmente ao sujeito da oração anterior (Mayol 2016, Teixeira *et al.* 2014). Com relação à produção de pronomes plenos, os dados do PB mantêm semelhanças e diferenças aos dados observados para o inglês.

A semelhança recai sobre o fato de a produção dessa expressão referencial não sofrer influência dos fatores semântico-pragmáticos que determinam sua interpretação. Tanto em inglês quanto em PB, o aspecto influenciou significativamente a interpretação pronominal, mas não suas taxas de produção. Esse dado corrobora as previsões do modelo de Kehler *et al.* (2008) ao evidenciar que produção e interpretação de pronomes estão sujeitos à influência de fatores distintos, indicando que podem se tratar de dois processos diferentes. Já as diferenças entre inglês em PB recaem sobre a influência da posição gramatical do antecedente na produção ou não do pronome pleno, algo que já era esperado devido às distinções tipológicas entre as línguas no que concerne a seus sistemas pronominais. Rohde (2008) relata forte influência de informações gramaticais na produção do pronome.

Em PB, vimos que a posição do antecedente não interferiu na produção do pronome pleno, que foi igualmente baixa tanto para retomada da fonte, na posição de sujeito, quanto na retomada do alvo, que ocupava a posição de objeto. Podemos concluir, portanto, que o valor do termo  $p(\text{pronome pleno}|\text{referente})$  é variável entre as línguas, possivelmente em decorrência de diferenças nas suas estruturas gramaticais.

Quanto à continuação das sentenças e interpretação do pronome pleno, nossas análises indicaram efeito principal de estímulo: comparativamente a textos com estímulo pronominal, textos de estímulo livre têm maiores chances de mencionar em sua sequência o alvo da sentença anterior. Esse efeito independe do aspecto verbal, pois dentre as sentenças de estímulo livre também há o mesmo efeito principal de aspecto observado para as sentenças de estímulo pronominal. Tais resultados nos trazem um dado que parece contraditório: ao mesmo tempo em que o pronome pleno não é usado com preferência para fazer menção à fonte ou alvo da sentença anterior, sua ocorrência gera mais interpretações de fonte do que nos contextos de estímulo livre, em que ele não ocorre. No entanto, esse dado é contraditório apenas se considerarmos que a produção de pronomes obedece aos mesmos critérios que a sua interpretação, o que não é o caso do modelo de Kehler *et al.* (2008).

Se retomarmos a equação em (3), expressa aqui novamente, a fim de explicar nossos resultados, podemos imaginar que tanto  $p(\text{pronome pleno}|\text{alvo})$  quanto  $p(\text{pronome pleno}|\text{fonte})$  em PB têm valores bastante próximos. De fato, o modelo de regressão logística ajustado indica que a probabilidade de uso de um pronome pleno para retomar o referente na posição de alvo é de 0,17, enquanto a probabilidade de uso da mesma expressão referencial como co-referencial à fonte é de 0,20, uma diferença que não foi identificada como significativa. Essa semelhança de valor entre os termos poderia explicar porque a interpretação do pronome, em PB, é mais suscetível a influências do termo  $p(\text{referente})$ , que informa a probabilidade de que um referente seja mencionado no contexto subsequente.

(3)

$$p(\text{referente}|\text{pronome pleno}) = \frac{p(\text{pronome pleno}|\text{referente}) p(\text{referente})}{p(\text{pronome pleno})}$$

Ainda assim, é possível que a probabilidade de mencionar o alvo dada a ocorrência de um pronome pleno, i.e.,  $p(\text{alvo}|\text{pronome pleno})$ , seja menor ou maior que a probabilidade de esse referente ser mencionado independente da expressão referencial escolhida, i.e.,  $p(\text{alvo})$ . Similarmente, o mesmo pode ocorrer com a probabilidade de interpretar um pronome pleno como a fonte da oração anterior. Isso ocorre pois o cálculo previsto para a interpretação do pronome leva em conta a interação desses dois termos e sua relação com  $p(\text{pronome pleno})$ , i.e., a probabilidade de uso de um pronome pleno.

Portanto, para uma avaliação satisfatória do modelo, é necessário que essas interações sejam consideradas de modo a comparar se a inserção das probabilidades obtidas em nossos modelos para os termos à direita da equação resultam em um  $p(\text{referente}|\text{pronome pleno})$  próximo aos valores observados em nosso experimento.

Em nosso estudo, as probabilidades obtidas através dos modelos de regressão logística apresentaram os valores dispostos na Tabela 3. Os resultados previstos pelo modelo de Kehler *et al.* e os observados em nosso experimento estão dispostos na Tabela 4.

Termo	probabilidade [intervalo de confiança - 95%]
p(pronome pleno)	0,18 [0,11 - 0,28]
p(fonte)	0,28 [0,17 - 0,44]
p(pronome pleno fonte)	0,20 [0,09 - 0,37]
p(alvo)	0,71 [0,55 - 0,82]
p(pronome pleno alvo)	0,17 [0,09 - 0,27]

**Tabela 3.** Probabilidades calculadas através do ajuste de modelos de regressão logística conforme seção de análise de resultados

Termo	Valor observado [intervalos de confiança]	Valor calculado pelo modelo de Kehler <i>et al.</i> a partir dos coeficientes registrados
p(fonte pronome pleno)	0,45 [0,35 - 0,56]	0,32
p(alvo pronome pleno)	0,54 [0,43 - 0,64]	0,66

**Tabela 4.** Probabilidades observadas e probabilidades previstas pelo modelo de Kehler *et al.* (2008) para a interpretação do pronome pleno ambíguo

Aplicando na equação em (3) os valores obtidos em nosso experimento, temos resultados para  $p(\text{fonte}|\text{pronome})$  e  $p(\text{alvo}|\text{pronome})$  similares aos obtidos pelo modelo de regressão logística a partir dos nossos dados. Ainda que os valores obtidos pelo modelo estejam próximos, mas não previstos, nos intervalos de confiança observados, eles se relacionam com  $p(\text{alvo})$  e  $p(\text{fonte})$  da mesma maneira observada nos dados do experimento. Percebemos que o uso de um pronome interage com  $p(\text{referente})$  de modo a diminuir o valor de  $p(\text{alvo}|\text{pronome})$  em comparação a  $p(\text{alvo})$  e de modo a aumentar  $p(\text{fonte}|\text{pronome})$  em comparação a  $p(\text{fonte})$ . Assim, há indicação de que o modelo de Kehler *et al.* seja capaz de prever variações em nossos dados que, a princípio, parecem contra-intuitivas.

Presumindo, portanto, que a interpretação do pronome pleno em PB, bem como em inglês, é resultado da interação entre um viés de produção e as expectativas geradas sobre qual referente tem mais chances de ser mencionado na sequência, podemos dar algumas respostas acerca da produção e interpretação do pronome pleno em PB. Se tivermos que responder se um pronome pleno tem tendência de ser usado para fazer referência ao sujeito da oração anterior, nossa resposta seria negativa, conforme atestam os dados de produção de pronomes plenos. Se tivermos que responder se um pronome pleno é preferencialmente interpretado como co-referente à fonte da sentença anterior, a resposta também é negativa, e corroborada pelos dados que mostram o efeito das relações de coerência na resolução pronominal.

Afinal, segundo nossos dados, a interpretação pronominal é em grande parte determinada pelo aspecto verbal, que seleciona relações de coerência específicas.

No entanto, se tivermos que responder se a presença de um pronome pleno atualiza as expectativas do próximo referente a ser mencionado, de modo a aumentar as chances de que haja uma menção ao sujeito (e fonte) da oração anterior, nossos dados suportam uma resposta afirmativa. Uma evidência adicional disso é a de que a presença de um pronome pleno muda a distribuição das relações de coerência de elaboração e consecutividade, que provavelmente são as responsáveis pelas diferenças observadas nas taxas de menção a fonte ou alvo entre as sentenças com estímulo pronominal e sem estímulo.

Essa dinâmica pode ser explicada uma vez que se assuma que pronomes plenos, em vez de preferência por um referente específico, introduzem vieses sobre como o discurso deve continuar. A pergunta a ser feita é o quanto de viés o pronome introduz na expectativa de que um determinado referente seja mencionado. Em inglês, esse viés coincide com uma grande preferência pelo uso do pronome pleno para fazer referência ao sujeito da oração anterior, o faz com que se indique o alto valor de  $p(\text{pronome}|\text{sujeito})$  como o responsável pelos resultados observados em Rohde (2008).

Em PB, esse viés também persiste, ainda que mais sutil. No entanto, sua ocorrência não deriva de uma preferência introduzida pelo uso do pronome pleno para retomar o sujeito da oração anterior, mas parece ser resultado da interação entre os três termos previstos no modelo de Kehler *et al.*

## 7. Conclusão

Este artigo se propôs a investigar como fatores gramaticais e informações pragmáticas atuam conjuntamente no estabelecimento da referência de um pronome pleno ambíguo, de modo a testar as predições do modelo de Kehler *et al.* (2008). Nossos dados corroboram o achado de que a produção e a interpretação de pronomes são melhores caracterizados como dois processos distintos. Com relação à interpretação dos pronomes plenos ambíguos, nossos resultados reforçam o achado já documentado para o inglês de que a resolução do pronome é em grande parte determinado pelo tipo de relação de coerência estabelecida entre as sentenças. Também corroboram a premissa do modelo testado de que o material linguístico encontrado pelo ouvinte/leitor tem papel importante na distribuição dessas relações de coerência. Contudo, diferentemente do encontrado em inglês, nossos resultados indicam que a produção do pronome pleno em PB não é influenciada pela posição gramatical de seu antecedente. Essa diferença, conforme argumentamos, seria a responsável pelas diferenças na interpretação de pronomes que podem ser observadas entre diferentes línguas.

Os dados que encontramos podem ser adequadamente explicados por um modelo que considera a interpretação de um pronome como resultado da interação entre um viés de produção e as expectativas geradas sobre o discurso. Esse modelo, que caracteriza a interpretação de um pronome por meio do Teorema de Bayes, é capaz de dar conta de resultados que à primeira vista parecem contra-intuitivos em PB.

Nesse sentido, os dados aqui reportados contribuem para validá-lo como uma proposta que consegue abarcar diferenças tipológicas entre línguas. Além disso, reforçam que não é possível investigar a resolução do pronome pleno como um processo que garante a retomada de determinada expressão referencial com base em preferências associadas a papéis temáticos e posição gramatical. Antes, a interpretação pronominal parece fazer parte de um processo inferencial mais geral de atualização de expectativas, em linha com pesquisas recentes que caracterizam o processamento linguístico como um processo ativo de antecipação de informações.

Cabe lembrarmos mais uma vez que nosso desenho experimental não distinguiu papel temático e posição gramatical. Estudos futuros devem se dedicar a investigar essa variável de confundimento, além de averiguar se o modelo aqui usado serve para explicar os dados de outros tipos de construções gramaticais em PB. Além disso, não sabemos de nenhum estudo que, assumindo o modelo de Kehler *et al.*, tenha se debruçado sobre a resolução de outros pronomes que não sejam o de terceira pessoa. Até o momento, essas são lacunas que ainda precisam ser preenchidas.

## Referências Bibliográficas

- Almor, Amit, Jefferson de Carvalho Maia, Maria Luiza Cunha Lima, Mirta Vernice, e Carlos Gelormini-Lezama. 2017. Language processing, acceptability, and statistical distribution: A study of null and overt subjects in Brazilian Portuguese, *Journal of Memory and Language* 92: 98-113.
- Alonso-Ovalle, Luis; Charles Clifton, Lyn Frazier e Susana Fernández-Solera. 2002. Null vs. Overt Pronouns and the topic-focus articulation in Spanish, *Journal of Italian Linguistics*, 14, 2: 151-169.
- Ariel, Mira. 1990. Accessing noun phrase antecedents, *Linguistics*, 96: 113-118
- Arnold, Jennifer E. 2001. The effects of thematic roles on pronoun use and frequency of reference continuation, *Discourse Processes*, 31: 137-162.
- Baayen, R. Harald; Douglas J. Davidson e Douglas M. Bates. 2008. Mixed-effects modelling with crossed random effects for subjects and items, *Journal of Memory and Language*, 59: 390-412.
- Barr, Dale. 2013. Coding categorical variables when analyzing factorial experiments with regression. [em linha]. Disponível em <http://talklab.psy.gla.ac.uk/tvw/catpred/>
- Barr, Dale, Roger Levy, Christoph Scheepers e Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal, *Journal of memory and language* 68, 3: 255-278
- Benjamini, Yoav e Yosef Hochberg. 1995. Controlling the False Discovering Rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B*, 57: 289-300.
- Brennan, Susan E., Marilyn W. Friedman e Carl J. Pollard. 1987. A centering approach to pronouns, em *Proceedings of the 25th meeting of the association for computational linguistics*, Stanford: 155-162.
- Carminati, Maria Nella. 2002. The processing of Italian subject pronouns, Tese de Doutorado. University of Massachusetts, Amherst, Inédita.

- Chambers, Craig G. e Ron Smyth. 1998. Structural parallelism and discourse coherence: A test of Centering Theory, *Journal of Memory and Language*, 39: 593-608.
- Dikker, Suzanne e Liina Pykkänen. 2013. Predicting language: MEG evidence for lexical preactivation, *Brain and Language*, 127, 1: 55-64.
- Filiaci, Francesca, Antonella Sorace e Manuel Carreiras. 2014. Anaphoric biases of null and overt subjects in Italian and Spanish: A cross-linguistic comparison, *Language and Cognitive Processes*, 29, 7: 825-843.
- Filiaci, Francesca. 2011. Anaphoric Preferences of Null and Overt Subjects in Italian and Spanish: a Cross-linguistic Comparison. Tese de Doutorado. University of Edinburgh, Edimburgo. Inédita.
- Fine, Alex B., Tim Florian Jaeger, Thomas A. Farmer e Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension, *PLoS ONE*, 8, 10: e77661.
- Fonseca, Maria Cristina M. e Erlândio Guerreiro. 2012. Resolução de correferência pronominal no português do Brasil, *Linguística*, 8, 2: 112-133.
- Fukumura, Kumiko e Roger P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference?, *Journal of Memory and Language*, 62: 52-66.
- Gordon, Peter C., Barbara J. Grosz e Laura A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse, *Cognitive Science*, 17: 311-347.
- Grosz, Barbara J., Aravind Joshi e Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse, *Computational Linguistics*, 21, 2: 202-225.
- Grüter, Theres, Hannah Rohde e Amyu Schafer. 2016. Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form, *Linguistic Approaches to Bilingualism*, 7, 2: 199-229.
- Hobbs, Jerry R. 1979. Coherence and coreference, *Cognitive Science*, 3, 1: 67-90.
- Kaiser, Elsi. 2013. Looking beyond personal pronouns and beyond English: Typological and computational complexity in reference resolution, *Theoretical Linguistics*, 39, 1-2: 109-122.
- Kehler, Andrew, Laura Kertz, Hannah Rohde e Jeffrey L. Elman. 2008. Coherence and Coreference Revisited, *Journal of Semantics*, 25, 1: 1-44.
- Kim, Kitaek, Theres Grüter e Amy J. Schafer. 2013. Effects of event-structure and topic/focus-marking on pronoun reference in Korean. Poster presented at the 26th Annual CUNY Conference on Human Sentence Processing, Columbia, SC.
- Kuperberg, Gina R. e T. Florian Jaeger. 2015. What do we mean by prediction in language comprehension?, *Language, Cognition and Neuroscience*, 31, 1: 32-50.
- Luegi, Paula. 2012. Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes. Tese de doutorado, Universidade de Lisboa. Inédito.

- Mayol, Laia. 2016. Asymmetries between interpretation and production in Catalan pronouns. Comunicacao apresentada no *7th Experimental Pragmatics Conference*. Köln. June 2017.
- Norcliff, Elisabeth, Alice C. Harris e T. Florian Jaeger. 2011. Cross-linguistic psycholinguistics and its critical role in theory development: early beginnings and recent advances, *Language and Cognition*, 8, 2: 167-205.
- Rohde, Hannah. 2008. Coherence-Driven Effects in Sentence and Discourse Processing. Tese de Doutorado. University of California, San Diego. Inédito
- Rohde, Hannah e William S. Horton, 2014. Anticipatory looks reveal expectations about discourse relations, *Cognition*, 133, 3: 667-691.
- Rohde, Hannah e Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production, *Language, Cognition and Neuroscience*, 29, 9: 912-227.
- Salverda, Anne P., Dave Kleinschmidte e Michael K. Tanenhaus. 2014. Immediate effects of anticipatory coarticulation in spoken-word recognition, *Journal of Memory and Cognition*, 1, 71: 145-163.
- Stevenson, Rosemary J., Rosalind A. Crawley e David Kleinman. 1994. Thematic roles, focusing and the representation of events, *Language and Cognitive Processes*, 9: 519-548.
- Teixeira, Elisângela Nogueira, Maria Cristina Micelli Fonseca, e Maria Elias Soares. 2014. Resolução do pronome nulo em Português Brasileiro: Evidência de movimentação ocular. *Veredas*, 18, 1: 281-301.
- Ueno, Mieko e Andrew Kehler. 2016. Grammatical and pragmatic factors in the interpretation of Japanese null and overt pronouns, *Linguistics*, 54, 6: 2057-2062.
- Wolf, Florian, Edward Gibson e Timothy Desmet. 2004. Discourse coherence and pronoun interpretation, *Language and Cognitive Processes*, 19: 665-75.