

Relações de coerência e resolução de anáforas

Neemias Silva de Souza Filho

Guilherme Luiz Andrade Santana da Silva¹

Vladimir Wanderley de Lima Rodrigues¹

Flávia Nayara Fernandes de Araújo²

Mahayana C. Godoy³

SOUZA FILHO, N.; SILVA, G.L.; RODRIGUES, V.W.; ARAÚJO, F.N.; GODOY, M.C. Relações de coerência e resolução de anáforas, *Linguística Rio*, vol.3, n.1, maio de 2017.

ISSN: 2358-6826

Informações do 1º autor
Graduando em Letras - Inglês na Universidade Federal do Rio Grande do Norte

Contato do 1º autor:
neemiassfilho@gmail.com

Outras informações
Enviado: 19 de janeiro de 2017
Aceito: 18 de abril de 2017
Online: 02 de junho de 2017

RESUMO: O objetivo deste trabalho é investigar se, em Português Brasileiro, as relações de coerência entre orações influenciam o processamento pronominal. Através de um experimento de leitura auto-cadenciada baseado no trabalho de Wolf *et al.* (2004), feito em inglês, investigamos a validade da proposta de Kehler (2002) de que a resolução anafórica está ligada à construção da coerência, e não ao uso de estratégias de natureza unicamente sintática. Não foram encontradas diferenças significativas que corroborassem algum dos modelos de resolução pronominal discutidos. Em nossa conclusão, explicamos esses resultados como motivados por possíveis diferenças tipológicas entre as línguas testadas e enfatizamos que o uso de replicações pode servir a um duplo propósito: ao mesmo tempo em que testa hipóteses da Psicolinguística para avaliar sua universalidade, beneficia a formação inicial do linguista ao colocá-lo em contato com trabalho de campo experimental ainda na graduação.

KEY WORDS: psicolinguística; anáfora; relações coesivas.

Introdução

Os estudos experimentais em psicologia cognitiva - e em psicolinguística, por extensão - têm passado por uma série de questionamentos quanto à confiabilidade de seus resultados (PASHLER e WAGENMAKERS, 2012). Dentre as práticas que levantam suspeita, está a baixa taxa de replicação de estudos (MAKEL, PLUCKER e HEGARTY, 2012), aliada a um número alto de análises estatísticas questionáveis (MASICAMPO e LALANDE, 2012) e à falta de clareza acerca de quais seriam as perguntas iniciais da pesquisa e quais seriam os resultados de análises *post-hoc* (VEER e GINER-SOROLLA, 2016). No campo da psicolinguística, adiciona-se a essas questões o fato de que diferenças tipológicas entre línguas deveriam despertar cautela quanto à alegação de universalidade dos resultados reportados para as poucas línguas investigadas (NORCLIFF, HARRIS e JAEGER, 2015). Nesse

¹ Graduando em Letras – Francês (UFRN).

² Licenciado em Letras – Francês (UFRN).

³ Doutora em Linguística (Unicamp), Professora do Departamento de Letras da UFRN.

sentido, a replicação de estudos em outras línguas serve não apenas para dar confiabilidade a dados já reportados, mas também para investigar como línguas tipologicamente tão distintas são processadas a partir dos mesmos substratos neurais.

Considerando, portanto, a necessidade de replicações de estudos prévios e a urgência de se investigar a suposta universalidade de achados sobre processamento linguístico, o presente trabalho tem como objetivo investigar se, em Português Brasileiro (PB), as relações de coerência entre orações impactam o processamento pronominal, a exemplo do que reportam Wolf *et al.* (2004) para o inglês. O estudo dos autores apresenta evidências empíricas de que a resolução anafórica se dá por meio da construção de relações de coerência entre sentenças, e não pelo uso de estratégias de natureza unicamente estruturais. Não é de nosso conhecimento que esse estudo tenha sido replicado em outras línguas, e os resultados em PB não corroboram os achados dos autores e nem as hipóteses levantadas por outros modelos de processamento de pronomes que também investigaram a língua inglesa. Conforme discutiremos, nossos dados apontam a necessidade de maiores estudos na área, especialmente considerando as divergências de mecanismos de processamento para línguas distintas reportadas nas últimas décadas (e.g., CUETOS, MITCHELL e CORLEY, 1996).

Uma vez que nossos resultados são fruto do trabalho final de um curso de graduação, discutimos brevemente a importância do trabalho aqui desenvolvido não apenas para o nosso campo de estudo, mas também para a formação inicial do linguista. Contudo, antes de partirmos para a descrição de nosso experimento, convém fazer um breve percurso sobre os modelos de processamento das anáforas pronominais.

1. Modelos de resolução pronominal

Quando utilizamos a língua em condições habituais, não nos deparamos com sentenças desconexas. As entidades sobre as quais se fala são introduzidas em um primeiro momento e, eventualmente, referenciadas novamente. Esse processo de retomada é responsável não só por manter os objetos mencionados previamente no discurso ativados na representação que construímos, mas também por criar as condições necessárias para a progressão referencial do texto. O

processo de referência é, portanto, de fundamental importância para que se possa construir uma representação discursiva coerente (TRAXLER, 2012).

Diversas são as maneiras por meio de que o processo de referência pode ser realizado, mas, no âmbito dos estudos de processamento da linguagem, é o uso de pronomes anafóricos que se provou como objeto de estudo mais prolífico. Mais especificamente, o estudo da resolução anafórica se deu, inicialmente, com trabalhos que exploravam o assinalamento de referentes a pronomes em contextos ambíguos (e.g., TYLER e MARSLEN-WILSON, 1982).

Um dos modelos de resolução de anáforas propostos postula uma preferência por ligar o pronome ao termo que ocupa a posição de sujeito da oração anterior (CRAWLEY, STEVENSON e KLEINMAN, 1990). Dessa maneira, comparando construções como (1a) e (1b)⁴, aquela deveria ser mais facilmente processada do que esta, hipótese que foi de fato corroborada por evidências empíricas (para uma revisão, ver WOLF *et al.*, 2004 e GODOY, 2010).

1. a) Joyce puniu Igor, e ela castigou Pedro.
b) Joyce puniu Igor, e ele castigou Pedro.

Em um experimento de leitura auto-cadenciada, entretanto, Chambers e Smyth (1998) encontraram evidências de que, em arranjos como (2), frases como (2b), em que o pronome referencia um antecedente em posição de objeto, são lidas mais rapidamente do que frases como (2a).

2. a) Joyce cumprimentou Igor, e Pedro parabenizou-a.
b) Joyce cumprimentou Igor, e Pedro parabenizou-o.

Esses resultados vão de encontro ao que é postulado por modelos de preferência pelo sujeito, o que levou os autores a propor um outro modelo de resolução anafórica para explicar seus dados, a *Parallel Preference*. Segundo essa proposta, haveria uma tendência a ligar o pronome a um referente previamente citado que ocupe a mesma posição sintática. Nesse sentido, a frase (2b) seria mais facilmente lida porque o pronome oblíquo pode ser associado a “Igor” - o que não

⁴ Os exemplos presentes neste texto, quando possível, são apresentados em português para facilitar a explicação da lógica dos experimentos citados. Assim, convém ressaltar que todos os estudos em questão foram realizados em inglês.

ocorre em (2a), dada a restrição imposta pelo gênero do clítico, que só permite o assinalamento de “Joyce” como antecedente.

Convém destacar que, além de explicar os resultados obtidos a partir das frases apresentadas em (2), o modelo da *Parallel Preference* também abrange os dados apresentados a favor do modelo de resolução pronominal que preveem uma preferência por ligar o pronome ao sujeito da oração anterior. Em (1a), por exemplo, a leitura mais rápida seria devido ao fato de que o pronome “ela” e o nome “Joyce”, único antecedente licenciado, ocupam a posição de sujeito em suas respectivas orações - o que não ocorre em (1b), visto que o único antecedente possível para o pronome “ele” é “Igor”, que ocupa a posição de objeto na oração anterior.

Os estudos sobre resolução anafórica são, entretanto, muito mais conflitantes do que poderíamos explicitar nesta seção. Kehler (2002) ressalta que pode haver divergência quanto à preferência de escolha de referentes pronominais mesmo entre frases com estruturas idênticas, mas com verbos de diferentes cargas semânticas e pragmáticas, constituindo uma situação que não pode ser explicada por nenhum dos modelos que discutimos até agora⁵. Kehler *et al.* (2008) criticam, nesse sentido, a pouca atenção reservada por esses modelos aos aspectos semânticos e pragmáticos da resolução pronominal, argumentando que esse processo seria guiado pelo estabelecimento das relações de coerência.

Uma consequência derivada dessa proposta é a de que as preferências de resolução de anáforas poderiam ser alteradas de acordo com a relação de coerência estabelecida entre as orações que contêm o antecedente e o pronome (WOLF *et al.*, 2004). Para testar essa hipótese, Wolf *et al.* (2004) criaram frases experimentais cujas orações estavam relacionadas ou por semelhança ou por causalidade, conforme os exemplos utilizados no estudo, apresentados em (3)⁶.

3. **a) Relação de semelhança, referência paralela:** *Fiona complimented Craig and similarly James congratulated him after the match but nobody took*

⁵ Para exemplificar situações desse tipo, Kehler (2002) cita exemplos de frases propostas por Winograd (1972), em que o referente pronominal licenciado não é determinado por traços gramaticais, mas sim semânticos e pragmáticos, como neste caso: “*The city council denied the demonstrators a permit because they feared/advocated violence*” (WINOGRAD, 1972, apud KEHLER, 2002, p. 146).

⁶ Os exemplos citados em (3) são apresentados por Wolf *et al.* (2004). Traduções nossas.

any notice. (Fiona elogiou Craig e, de maneira semelhante, James parabenizou-o depois do jogo, mas ninguém notou.)

b) Relação de semelhança, referência não paralela: *Fiona complimented Craig and similarly James congratulated her after the match but nobody took any notice.* (Fiona elogiou Craig e, de maneira semelhante, James parabenizou-a depois do jogo, mas ninguém notou.)

c) Relação de causalidade, referência paralela: *Fiona defeated Craig and so James congratulated him after the match but nobody took any notice.* (Fiona derrotou Craig e, por isso, James parabenizou-o depois do jogo, mas ninguém notou.)

d) Relação de causalidade, referência não paralela: *Fiona defeated Craig and so James congratulated her after the match but nobody took any notice.* (Fiona derrotou Craig e, por isso, James parabenizou-a depois do jogo, mas ninguém notou.)

Através de uma atividade de leitura auto-cadenciada, os pesquisadores analisaram o tempo de leitura de um pronome anafórico que sempre ocupava a posição de objeto e que poderia ter como antecedente licenciado o sujeito (configurando referência não paralela, nos termos da *Parallel Preference*) ou o objeto (referência paralela) da oração anterior. Para sentenças com relação de semelhança, manteve-se a preferência pelo pronome ligar-se ao objeto; para sentenças em relação de causa-efeito, o pronome era lido mais rapidamente quando retomava o sujeito. Assim, os resultados obtidos pelos autores indicaram que a atribuição de referentes não apresentava a mesma tendência em frases com relações de coerência distintas, corroborando a proposta de Kehler (2002).

Frente a esses resultados e tendo em vista que as teorias discutidas são baseadas majoritariamente em estudos feitos com a língua inglesa, adaptamos o desenho experimental de Wolf *et al.* (2004) para fazer uso de pronomes clíticos na posição de objeto. Os detalhes metodológicos serão discutidos com minúcia na seção apropriada, mas convém antecipar que, assim como é feito no trabalho que tomamos por base, utilizamos um experimento de leitura auto-cadenciada e lançamos mão de manipulações que visavam estabelecer relações de semelhança ou causa-efeito entre orações.

2. Metodologia

Nosso experimento foi desenvolvido com base no estudo de Wolf *et al.* (2004), com o objetivo de testar a hipótese dos autores em Português Brasileiro. Para tanto, criamos 16 frases experimentais compostas por 3 orações cada. As relações de coerência e as relações de correferência pronominal entre a primeira e a segunda oração de cada frase experimental foram utilizadas como variáveis independentes, cada uma com 2 níveis. As relações de coerência poderiam ser de semelhança ou de causa-efeito. As referências pronominais poderiam ser paralelas ou não paralelas⁷. Dessa forma, em um design 2x2, foram estabelecidas 4 condições por frase (cf. Tabela 1).

	Condição	Frase experimental
1.	<u>Relação de semelhança e Referência paralela</u>	Graça elogiou Fábio <u>e da mesma forma</u> Luiz felicitou- o com gosto depois da partida mas ninguém deu muita importância ao fato.
2.	<u>Relação de semelhança e Referência não paralela</u>	Graça elogiou Fábio <u>e da mesma forma</u> Luiz felicitou- a com gosto depois da partida mas ninguém deu muita importância ao fato.
3.	<u>Relação de causa-efeito e Referência paralela</u>	Graça derrotou Fábio <u>e por causa disso</u> Luiz felicitou- o com gosto depois da partida mas ninguém deu muita importância ao fato.
4.	<u>Relação de causa-efeito e Referência não paralela</u>	Graça derrotou Fábio <u>e por causa disso</u> Luiz felicitou- a com gosto depois da partida mas ninguém deu muita importância ao fato.

Quadro 1: Descrição das condições e itens experimentais

Como variável dependente, foi considerado o tempo de leitura do verbo da segunda oração, ao qual está ligado o pronome clítico. Nossa hipótese, de acordo com o que é previsto pelo modelo de resolução pronominal proposto por Kehler (2002), era de que menores tempos de leitura seriam registrados nas condições 1

⁷ O paralelismo da referência pronominal, como pode ser observado na Tabela 1, refere-se à concordância ou divergência entre as funções sintáticas do pronome e de seu referente (i.e., se ambos são objeto, a referência é paralela; se um é sujeito e o outro é objeto, a referência é não paralela).

e 4⁸. Para controlar possíveis efeitos de *spill-over*, a região posterior à crítica era sempre preenchida por locuções adverbiais de mesma extensão, seguidas por uma terceira oração. Levamos em consideração a quantidade de sílabas dos nomes próprios de cada frase, cada um com duas sílabas. Controlamos o tamanho e o tempo dos verbos, que sempre tinham 4 sílabas e estavam flexionados no pretérito perfeito.

Além das frases experimentais, foram criadas mais 32 frases distratoras. As sentenças foram distribuídas em 4 listas. Cada lista recebeu as 16 frases (4 frases de cada condição) e as 32 distratoras. Os itens experimentais foram distribuídos de modo que nenhum sujeito visse mais de uma condição de cada frase. O *script* dos procedimentos foi preparado e executado por meio do programa *Psychopy*.

Os participantes foram orientados sobre os procedimentos necessários à execução do experimento, que utilizou a técnica experimental de leitura auto-cadenciada não-cumulativa, ou seja, as palavras apareciam uma a uma à medida que o sujeito pressionasse a barra de espaço. Após a leitura de cada frase, os sujeitos respondiam sim ou não a uma questão de compreensão, para garantir que estivessem atentos à tarefa.

Contamos com um total de doze participantes no experimento, em sua maioria da comunidade acadêmica da Universidade Federal do Rio Grande do Norte. Todos eram maiores de idade, falantes do Português como língua materna e desconhecedores dos objetivos reais do estudo. Cada sujeito levou em média vinte minutos para a realização do experimento.

3. Resultados

Foram descartados os dados de um participante que errou mais de 20% das questões de compreensão. O restante dos dados foi analisado. Como a distribuição observada dos dados não foi normal, a análise estatística deu-se por meio do teste *Mann-Whitney*, a fim de verificar se as amostras das condições analisadas seriam provenientes da mesma população. Não houve efeito principal de coerência ($p = 0.69$, $W = 2964$) ou paralelismo ($p = 0.780$, $W = 3000$) para os tempos de leitura do

⁸ Para uma discussão a respeito das previsões feitas por cada modelo de resolução anafórica mencionado neste trabalho, ver a seção “*Predictions*” de Wolf *et al.* (2004).

trecho crítico que continha verbo+pronome. As comparações múltiplas⁹ dos tempos de leitura de cada uma das condições também não indicou qualquer efeito de interação entre as variáveis controladas ($p > 0,05$). Também não houve qualquer efeito registrado na região posterior à região crítica.

4. Discussão

Conforme discutido anteriormente, o resultado reportado por Wolf *et al.* (2004) indica que o processamento pronominal *online* não pode ser explicado por heurísticas desenvolvidas a partir de pistas meramente estruturais. Segundo os autores, a resolução pronominal seria um subproduto de processos cognitivos relacionados ao estabelecimento das relações de coerência entre sentenças. Nossos resultados, contudo, não corroboraram o achado dos autores.

Outro dado relevante que se pode extrair de nosso estudo é o fato de os resultados aqui reportados também não corroborarem modelos estruturais de resolução pronominal. Segundo a hipótese da preferência pelo sujeito, a região crítica das condições não-paralelas seria lida mais rapidamente, o que não ocorreu. Por outro lado, caso a preferência por manter um paralelismo estrutural fosse a estratégia utilizada para resolver expressões pronominais, teríamos observado menores tempos de leitura para a região crítica nas condições em que havia paralelismo estrutural entre as duas orações. Apesar de todas essas previsões, os tempos de leitura para as quatro condições foi essencialmente o mesmo, sem registro de qualquer efeito principal ou de interação das variáveis controladas.

Os resultados encontrados podem ter duas explicações alternativas. Em primeiro lugar, há de se considerar as fragilidades do experimento aqui descrito, notadamente o baixo número de participantes. Enquanto Wolf *et al.* (2004) contaram com a participação de 40 voluntários em seu experimento, nosso estudo, principalmente por se tratar de um trabalho desenvolvido em uma disciplina de graduação, partiu da análise de dados de apenas 11 participantes.

Apesar da ressalva feita quanto a essa fragilidade em nosso estudo, os dados reportados, por indicarem que pode não haver efeito semelhante em PB, apontam a necessidade de se investigar a questão com mais profundidade. Trabalhos recentes alertam para o fato de os estudos em psicolinguística se

⁹ Como foram realizadas comparações múltiplas, os ajustes dos p-valores foram feitos utilizando o critério de *false discovery rate*.

ocuparem de apenas 0,6% das línguas do mundo (NORCLIFF, HARRIS e JAEGER, 2015), o que impossibilitaria assumir a universalidade dos mecanismos de processamento documentados. O questionamento da universalidade dos achados em psicolinguística se torna ainda mais urgente se considerarmos as diferenças documentadas entre os mecanismos de processamento de línguas distintas (CUETOS, MITCHELL e CORLEY, 1996; TURNBULL *et al.*, 2015; PHAM e BAAYEN, 2015). Nesse contexto, o fato de nossos resultados não espelharem o que foi reportado pela língua inglesa pode assinalar mais uma dessas diferenças.

Por fim, há ainda um comentário a ser feito sobre o experimento descrito que não diz respeito, exatamente, aos modelos de processamento pronominal. Para além da contribuição teórica, acreditamos que a replicação de experimentos em cursos de graduação cumpre uma dupla função. Por um lado, contribui para a formação o futuro linguista ao engajá-lo com pesquisa científica atual. Esse engajamento não apenas aproxima o aluno de um tópico de pesquisa específico, mas o familiariza com o método científico e com técnicas experimentais. Além disso, conforme argumentamos na introdução deste trabalho, as replicações são importantes para o desenvolvimento das pesquisas experimentais, pois ora corroboram estudos prévios, ora desafiam seus resultados. Nesse último caso, como acontece com o presente estudo, essas replicações abrem a possibilidade de revisitar velhas questões para avançar o que sabemos sobre os mecanismos de processamento linguístico.

REFERÊNCIAS

- CHAMBERS, G. C.; SMYTH, R. Structural parallelism and discourse coherence: a test of Centering Theory. *Journal of Memory and Language*, n. 39, p. 593-608, 1998.
- CRAWLEY, R., STEVENSON, R., & KLEINMAN, D. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, n. 4, p. 245- 264, 1990.
- CUETOS, F.; MITCHELL, D. C.; CORLEY, M. M. B. Parsing in different languages. In: CARREIRAS, M; GARCIA-ALBEA, J. E.; SEBASTIAN-GALLES, N. (Orgs.). *Language processing in Spanish*. Hillsdale, NJ: Erlbaum, 1996. p. 145-187.
- WOLF, F.; GIBSON, E.; DESMET, T. Discourse coherence and pronoun resolution. *Language and Cognitive Processes*, v. 9, n. 6, p. 665-675, 2004.
- GODOY, M. C. *Resolvendo a anáfora conceitual: um olhar para além da relação antecedente/anafórico*. 2010. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas. 2010.

KEHLER, A. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications, 2002.

KEHLER, A. et al. Coherence and coreference revisited. *Journal of Semantics*, v. 25, n. 1, p. 1-44, 2008.

MAKEL, M. C.; PLUCKER, J. A.; HEGARTY, B. Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science*, v.7, n. 6, p. 537-542, 2012.

MASICAMPO, E. J.; LALANDE, D. R. A peculiar prevalence of p values just below .05. *The Quarterly Journal of Experimental Psychology*, v. 65, n. 11, p. 2271-2279, 2012.

NORCLIFFE, E; HARRIS, A. C.; JAEGER, F. Crosslinguistic psycholinguistics and its critical role in theory development: early beginnings and recent advances. *Language, Cognition and Neuroscience*, v. 30, n. 9, p. 1009-1032, 2015.

PASHLER, H.; WAGENMAKERS, E. J. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*, v. 7, n. 6, p. 528-530, 2012.

PEIRCE, J. W. PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, v. 162, n. 1, p. 8-13, 2007.

PHAM, H.; BAAYEN, H. Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition and Neuroscience*, v. 30, n. 9, p. 1077-1095, 2015.

TRAXLER, M. J. *Introduction to Psycholinguistics: understanding language science*. 1. ed. Malden, MA: Wiley-Blackwell, 2012.

TURNBULL, R.; BURDIN, R.S.; CLOPPER, C.G.; TONHAUSER, J. Contextual predictability and the prosodic realization of focus: a cross-linguistic comparison. *Language, Cognition and Neuroscience*, v. 30, n. 9, p. 1061-1076, 2015.

ABSTRACT: This paper intends to investigate, in Brazilian Portuguese, whether coherence relations between clauses can influence pronominal resolution. Using a self-paced reading experiment inspired by Wolf *et al.* (2004), we explore the validity of Kehler's (2002) proposal that anaphora resolution is linked to coherence building, and not to the use of syntax-based strategies. Results showed no significant differences that could support neither the coherence-based model nor previously developed syntax-based resolution models. In our conclusion, we interpret these results as motivated by possible typological differences between the studied languages. Furthermore, we emphasize that replications could serve a dual-purpose: testing psycholinguistic hypotheses to verify their universality whilst providing an opportunity for linguists to familiarize themselves with experimental research early on in their education.

KEYWORDS: psycholinguistics; anaphora; cohesive relations.

SOUZA FILHO, N.; SILVA, G.L.; RODRIGUES, V.W.; ARAÚJO, F.N.; GODOY, M.C. Relações de coerência e resolução de anáforas, *Linguística Rio*, vol.3, n.1, maio de 2017.

ISSN: 2358-6826

Outras informações

Enviado: 19 de janeiro de 2017

Aceito: 18 de abril de 2017

Online: 02 de junho de 2017

