

## COGNIÇÃO E PROCESSAMENTO LINGUÍSTICO: O CASO DAS RELATIVAS DE ENCAIXAMENTO CENTRAL E RECURSIVO

Cognition and language processing:  
the central and recursive embedded relative clauses cause

Eduardo Kenedy (UFF/CNPq)

### Resumo

Neste texto, são apresentados resultados experimentais de pesquisa acerca do encaixamento central e recursivo de orações relativas. Relativas desse tipo são intrigantes porque, diferentemente do que se passa com as relativas de encaixe recursivo à direita, o encaixe central recursivo é linguisticamente improcessável em qualquer língua natural. As razões para essa impossibilidade psicolinguística são ainda alvo de debate na literatura especializada (ver, por exemplo, INGVE, 1960; MILLER & CHOMSKY, 1963; MACDONALD ET AL., 1994; GIBSON, 2001 e GORDON et al., 2001, para uma perspectiva histórica). Os dados experimentais *off-line* e *on-line* aqui apresentados e discutidos convergem em favor da hipótese de que o processamento psicolinguístico – e a consequente sensação de (in)aceitabilidade – de relativas encaixadas ao centro, sendo seu encaixe único ou recursivo, é fortemente influenciado por fatores extralinguísticos, notadamente, o custo de manutenção na memória de diferentes entidades do discurso (SNs plenos, nomes próprios, pronomes dêiticos). Tais achados compõem evidência de que a inaceitabilidade universal das relativas de encaixe central recursivo provavelmente decorra das restrições que memória de trabalho humana impõe ao *parser*, e não de fatores de natureza puramente linguística.

**Palavras-chave:** Encaixamento de orações, orações relativas, recursividade, encaixe central.

### Introdução

**N**a literatura linguística, as “orações subordinadas adjetivas restritivas”, também denominadas “cláusulas relativas” ou, simplesmente, “orações relativas” são um dos temas mais investigados. Em especial, um fenômeno relacionado a um emprego específico de relativas vem intrigando os linguistas desde, pelo menos, os anos 1960: as orações relativas de encaixe central recursivo (cf. INGVE, 1960; MILLER & CHOMSKY, 1963; entre outros). Esse tipo de estrutura sintática é gerado por uma computação linguística que consiste na introdução de uma nova oração relativa (OR) no domínio de outra OR previamente apresentada na linearidade da sentença. Em (1a), a seguir, apresenta-se uma OR de encaixe central único. Em (1b), por sua vez, vê-se uma nova OR encaixada no interior da relativa presente em (1a), fato que exemplifica o encaixe central recursivo de ORs.

(1a) O rato [ORque o gato perseguia] fugiu.

(1b) O rato [ORque o gato [ORque o cachorro espantou] perseguia] fugiu.

As gramáticas das línguas naturais não fazem restrições à quantidade de ORs que podem ser encaixadas recursivamente ao centro de uma sentença. Com efeito, o encaixamento central de duas ou mais ORs é um expediente tão legitimamente gramatical quanto o encaixe ao centro de uma única OR. No entanto, o encaixe central de uma segunda OR, tal como ocorre em (1b), torna a estrutura resultante inaceitável para a maioria dos falantes das línguas. Na verdade, conforme se ilustra em (2), a introdução de uma terceira OR, no domínio da segunda previamente existente na linearidade do estímulo linguístico, resulta numa estrutura sintática que, apesar de gramatical, é inaceitável para virtualmente qualquer ser humano.

1

(2) O rato [OR que o gato [OR que o cachorro [OR que o menino espantou] mordeu] perseguia] fugiu.

Os exemplos (1) e (2) estão em língua portuguesa, porém o mesmo fenômeno se registra em todas as línguas naturais conhecidas, independentemente de sua tipologia morfossintática. Numa língua núcleo final como, por exemplo, no japonês, o nome que sofre a relativização é linearizado depois da OR. Nessas línguas, mais de uma OR encaixada ao centro leva inevitavelmente à inaceitabilidade da sentença. Interessantemente, a inaceitabilidade de múltiplas relativas encaixadas ao centro ilustra um dos poucos fenômenos linguísticos sem exceção nos estudos sobre as diferentes línguas naturais.

Miller & Chomsky (1963) chamaram a atenção para o fato de que ORs das dos tipos em (1b) e (2) não violam qualquer aspecto puramente linguístico da competência linguística de um falante de uma língua humana. Antes, a inaceitação desse tipo de estrutura é um fenômeno da ordem do desempenho linguístico, para o qual intervêm

---

<sup>1</sup> Gibson (2001) indica haver um percentual muito pequeno de falantes de inglês que conseguem processar *on-line* até duas orações relativas de encaixa central, índice que é reduzido a zero com a introdução de uma terceira relativa encaixada ao centro.

fatores que até o presente não foram suficientemente esclarecidos nem pela teoria da gramática, tampouco pela psicolinguística.

Ora, ao se assumir a clássica dicotomia chomskiana (1965), o estudo das relativas encaixadas recursivamente ao centro fará emergir o seguinte questionamento: se aspectos do conhecimento linguístico não vêm ao caso, quais seriam os fatores da performance que provocam a rejeição universal desse tipo de estrutura sintática? O presente texto tem como objetivo justamente fornecer evidências experimentais que buscam encontrar respostas, ainda que parciais, para tal problema. Um conjunto de três experimentos explorará a noção sintática de *localidade* a fim de verificar a hipótese de que o processamento psicolinguístico de ORs, seja com encaixe único ou recursivo, pode ser variável conforme a natureza referencial das expressões nominais que separam o alvo da relativa de seu respectivo predicador no domínio da oração matriz. Tal hipótese foi motivada por evidências de natureza intuitiva que indicam, por exemplo, que a presença de um pronome pessoal dêitico no interior da segunda relativa torna o processamento de duas ORs encaixadas ao centro consideravelmente menos custoso. Compare-se o efeito subjetivo de maior inaceitabilidade em (3a) por oposição a (3b) a seguir.

(3a) O rato [OR que o gato [OR que o cachorro espantou] perseguia] fugiu.

(3b) O rato [OR que o gato [OR que ele espantou] perseguia] fugiu.

Em (3a), a expressão nominal presente na segunda relativa é um sintagma de valor referencial pleno (“cachorro”), que contrasta com natureza dêitica, referencialmente menos custosa (cf. BENVENISTE, 1976), do pronome sujeito da segunda OR, em (3b). Essa diferença de status referencial entre nomes plenos e pronomes pessoais poderia explicar a sensação subjetiva que licencia (3b), mas rejeita (3a). A hipótese a ser aqui defendida explorará essa intuição: no cálculo do custo da localidade no processamento de ORs de encaixe central recursivo deve-se considerar a natureza referencial de elemento interveniente entre o alvo da primeira OR e seu respectivo predicador, linearizado após a segunda OR.

O presente texto está organizado em cinco seções, seguidas de um sumário dos resultados da pesquisa. Na seção 1, descreve-se a hipótese chomskiana, dos anos 60, que explica a inaceitabilidade de ORs centrais recursivas em função de parâmetros puramente sintáticos – o número de relações de dependências estruturais não saturadas por algum

item lexical. Na seção 2, apresenta-se a noção de *entidade referencial* (cf. GIBSON, 2001 e GORDON et al., 2001), a ser considerada como uma das métricas fundamentais para o cômputo da localidade no processamento das ORs centrais. Na seção seguinte, descreve-se um experimento de julgamento imediato de aceitabilidade, que verificou como 50 brasileiros adultos perceberam relativas com dois encaixes centrais que variavam de acordo com a natureza do nome presente na segunda relativa (se um referencial pleno, um nome próprio ou um pronome dêitico). A seção 4 apresenta e comenta um experimento de escolha de paráfrase, em que 40 crianças do primeiro ciclo do ensino fundamental escolheram, entre duas opções, a sentença sem OR que parafraseava uma sentença com uma OR de encaixe central único. Um experimento de leitura segmentada autocadenciada é discutido na seção 5. Nele, 20 brasileiros adultos leram sentenças com dois tipos de ORs de encaixe único ao centro, que se diferenciavam pela ocorrência das mesmas variáveis controladas no experimento de julgamento imediato de aceitabilidade. Como conclusão, na seção “Resultados” faz-se uma análise geral dos possíveis achados aduzidos nas seções anteriores e apresenta-se uma reflexão sobre os fatores do desempenho linguístico que podem ser responsáveis pela improcessabilidade – e a conseqüente rejeição universal – de ORs de encaixe central recursivo.

### **1. Pressupostos teórico-metodológicos: a explicação sintática**

Desde Miller & Chomsky (1963), a hipótese dominante na teoria linguística atribuiu a inaceitabilidade universal de ORs de encaixe central recursivo à existência, nesses tipos de OR, de duas ou mais estruturas de dependência em longa distância não saturadas por algum item lexical. De fato, uma relativa de encaixe central único possui a característica de retardar a saturação da predicação referente à expressão nominal, no domínio da oração matriz, que é modificada pela OR. Assim, em (1a), a predicação de “O rato” só é estabelecida quando da introdução do verbo “fugiu”, linearizado após a OR, a qual, por sua vez, introduz na sentença uma nova expressão referencial – “o gato” – e a satura imediatamente, com a linearização local de seu predicador “persegue”. Nesse tipo de estrutura sintática, gera-se, portanto, uma e somente uma relação de dependência de longa distância (o alvo da relativa e seu respectivo predicador não local). Por contraste às ORs de encaixe central, ORs encaixadas à direita (ou à esquerda, nas línguas do tipo núcleo final) modificam expressões nominais cuja predicação é saturada imediatamente

nos limites da oração matriz, tal como se ilustra em (4). Nesse exemplo, “o rato” já tem sua predicação estabelecida pelo verbo “persegue” antes mesmo da introdução da OR, não gerando, dessa forma, relações sintáticas não locais.

(4) O gato perseguia o rato [OR que fugiu].

É em função dessa saturação local que ORs encaixadas à direita não apresentam restrições a possíveis encaixamentos recursivos, tal como explorado no famoso poema “Quadrilha”, de Carlos Drummond de Andrade. No caso, ilustrado em (5), todas as expressões nominais (nomes próprios) são saturadas localmente como complemento do verbo que as antecede.

(5) João amava Teresa [OR que amava Raimundo [OR que amava Maria [OR que amava Joaquim [OR que amava Lili [OR que não amava ninguém]]]]].

ORs de encaixe central recursivo – também denominadas *aninhadas* ou, em inglês, *nested* – fazem crescer exponencialmente a complexidade na saturação não local da predicação existente em OR centrais de encaixe único – e essa seria, na hipótese chomskiana, a causa da dificuldade de processamento desse tipo de estrutura sintática. Ou seja, o encaixe de uma segunda OR ao centro, como ocorre em (1b), deixaria não somente a predicação de “O rato” em suspenso, mas também a de “o gato”. Dessa forma, esse tipo de estrutura geraria pelo menos duas relações de dependência sintática não local. Veja-se, para ilustração, o que acontece em (6a), exemplo em que há duas ORs encaixadas ao centro. A predicação de “O rato” e “o gato” já estão em aberto (não saturadas) no momento em que mais uma expressão nominal (“o cachorro”) é introduzida na sentença. A predicação desse último constituinte é estabelecida imediatamente pelo verbo “espantou” e, só então, a predicação de “o gato” é saturada com a apresentação do item “persegue” (primeira relação de longa distância). Logo em seguida, o predador de “O rato” (o verbo “fugiu”) é finalmente introduzido (segunda relação de longa distância) e o processamento da sentença pode ser concluído. Note-se, por contraste, que, nas relativas de encaixe à direita, esse tipo de complexidade não acontece, afinal, como dito, a introdução de uma OR ocorre somente quando a predicação da expressão nominal

respectiva já se encontra saturada localmente (cf. 6b). Essa é a essência da explicação clássica chomskiana.

(6a) O rato [OR que o gato [OR que o cachorro espantou] perseguia] fugiu].

(6b) O cachorro espantou o gato [OR que perseguia o rato] [OR que fugiu].

A explicação puramente sintática para o fenômeno das improcessabilidade psicolinguística e a conseqüente rejeição universal de duas ou mais ORs encaixadas ao centro não conseguiria explicar a diferença de aceitabilidade entre (3a) e (3b), citados mais acima. Deve-se notar que, nesse par de sentenças, o número de dependências de longa distância existentes é rigorosamente o mesmo. Com efeito, para além da oposição entre o pronome “ele” e a expressão referencial plena “cachorro”, (3a) e (3b) são sentenças exatamente idênticas. Por que, então, não seriam ambas rejeitadas da mesma forma? Na tentativa de superar as limitações da explicação clássica, será apresentada a seguir uma abordagem que leva em consideração a natureza referencial da expressão nominal existente no domínio das ORs e seu respectivo impacto no processamento psicolinguístico.

## **2. Pressupostos teórico-metodológicos: a natureza referencial das expressões nominais**

Como visto, de acordo com a concepção dominante em teoria da gramática, a razão para a improcessabilidade de ORs dupla ou triplamente aninhadas seria puramente estrutural: a ocorrência de duas ou mais relações sintáticas não locais impõe um custo de processamento que a memória humana simplesmente não consegue suportar (cf. YNGVE, 1960; MILLER & CHOMSKY, 1963, entre outros). No entanto, desde o início dos anos 70, estudos psicolinguísticos independentes vêm indicando que a razão para o alto custo demandado por estruturas dessa natureza é possivelmente derivado de outros fatores para além de questões de ordem sintática. Por exemplo, Kimball (1973), em seu clássico “Seven principles of surface structure parsing in natural language” (“Sete princípios de parsing sobre estruturas de superfície em língua natural”), anunciava, com o Princípio 4, denominado “Two sentences” (“Duas frases”), que o *parsing* em línguas naturais não é capaz de analisar mais do que duas orações simultaneamente. Para o autor, seu princípio apenas atestaria um fato relativo à natureza do processamento sintático. A

causa mais profunda para esse fato deveria ser revelada pela pesquisa psicolinguística então programática. Anos depois, MacDonald et al. (1994) indicaram, de maneira explícita, diversos tipos de informação extrassintática que podem ser levadas em conta pelo *parser* durante o processamento de sentenças estruturalmente ambíguas ou sem ambiguidade, abrindo assim espaço para a exploração teórica e experimental acerca do papel que fatores lexicais e discursivos desempenham na computação mental de frases em tempo real. É na esteira dessas trabalhos que surgiram os estudos de Gibson (2001) e Gordon et al. (2001), que se caracterizam como os primeiros em que a questão das relativas de encaixe central recursivo foi especificamente contemplada.

Gibson (2001) e Gordon et al. (2001) investigaram experimentalmente dados intuitivos formulados inicialmente por Bever (1970), segundo os quais a natureza da expressão nominal contida no domínio da segunda (da terceira etc.) OR pode tornar esse tipo de estrutura sintática mais facilmente processável. De fato, comparem-se os exemplos (7a-b) do inglês, estruturalmente idênticos aos apresentados em (3) do português. As relações sintáticas de longa distância são precisamente as mesmas em ambas as sentenças, entretanto o custo de processamento de (7b) parecer ser muito menor do que o de (7a), do que resultaria sua maior aceitabilidade.

(7a) The reporter [ORwho the senator [OR that John met] attacked] disliked the editor.

O repórter que o senador que João encontrou atacou desagradou ao editor.

(7b) The reporter [OR who everyone [OR that I met] attacked] disliked the editor.

O repórter que todo mundo que eu encontrei atacou desagradou ao editor.

Para dar conta da possível assimetria no processamento e na percepção de estruturas como (7a) x (7b), Gibson (2001) explorou a possibilidade de o cálculo do processamento de ORs de encaixe central recursivo ser medido por meio do custo de manutenção, na memória de trabalho, das entidades referenciais à espera de sua respectiva predicação. Assim, cada unidade desse tipo imporia custo para memória (medido em uma unidade para cada entidade referencial) quando sua predicação não é saturada localmente, por oposição a expressões não referenciais (por exemplo, pronomes pessoais dêiticos), cujo custo para o processamento seria igual a zero.

A teoria formulada por Gibson mostrou-se interessante, inclusive, para explicar a assimetria encontrada, em seus experimentos, entre o processamento de relativas de sujeito e de objeto com encaixe recursivo único. Relativas de sujeito demonstraram menores latências num experimento de leitura segmentada autocadenciada, por contraste a tempos de reação maiores na leitura de relativas de objeto (GIBSON, 2001, p.111-113). Para o autor, isso ocorreu porque, nas relativas de sujeito, nenhuma entidade referencial no domínio da relativa tem sua predicação estabelecida à longa distância, dado que o verbo da relativa satura localmente seu wh-sujeito e seu objeto. Já nas relativas de objeto, o wh-objeto é separado de seu predicador por uma entidade referencial interveniente (a expressão nominal sujeito da relativa), o que torna sua saturação não local. Para Gibson, é em função dessa assimetria imposta na representação dos dois tipos de estrutura que os tempos de leitura são maiores para estruturas como (8a) – relativas de objeto –, por oposição a relativas de sujeito, como em (8b).

(8a) The reporter [ORwho the photographer sent \_ to the editor] hoped for a story.  
O repórter que o fotógrafo enviou para o editor esperava uma história.

(8b) The reporter [ORwho \_ sent the photographer to the editor] hoped for a story.  
O repórter que enviou o fotógrafo para o editor esperava uma história.

No caso específico das ORs encaixadas ao centro, Gibson (2001) indicou que a presença, no domínio da OR, de entidades menos custosas referencialmente do que a expressão nominal plena alvo da relativização poderá diminuir o custo de processamento desse tipo de estrutura. O exemplo máximo dessa “facilitação” são os casos de relativas com pronomes pessoais dêiticos na função de sujeito, como ilustrado em (3b) e (7b). Pronomes desse tipo seriam, de acordo com Gibson, “gratuitos” para a memória de trabalho usada durante o processamento linguístico, no sentido de que apontam para as pessoas do discurso (e, assim, são não referenciais ou são apenas indiretamente referenciais) e não impõem custos para a memória. O extremo oposto dos pronomes pessoais dêiticos seriam as expressões nominais (SNs) plenas indefinidas, como “um senador”, “uma casa”, que seriam as representações mais custosas para o processamento, já que introduzem entidades plenamente referenciais novas no discurso, que precisam ser ativadas e armazenadas na memória de trabalho. Entidades referências plenas definidas



(“o senador”, “a casa”) e, por isso, velhas no discurso – isto é, conhecidas pelo locutor e seu interlocutor – seriam menos custosas do que expressões indefinidas, visto que se encontram ativas na memória durante a construção mental do discurso. Nomes próprios, por sua vez, seriam menos custosos do que expressões nominais plenas, definidas ou não, dado que não demandam o tipo de representação plena que uma expressão nominal impõe para a memória de longo prazo (recuperação de um item lexical) e para a memória de trabalho (reativação e manutenção de um item lexical). O *continuum* da complexidade referencial proposta por Gibson (2001) é representado na ilustração que se segue.

/

SNs indefinidos > SNs definidos > nomes próprios > pronomes pessoais

**Ilustração 1. Do mais complexo ao menos complexo. Representações de entidades na memória.**

Gordon et al. (2001) apresentam um modelo psicolinguístico que faz as mesmas previsões que o paradigma acima. A diferença é que se trata de um modelo menos explícito, o qual faz alusão somente à maior ou à menor “diferença” entre as entidades referenciais que devem ser mantidas na memória de trabalho durante o processamento de ORs encaixadas ao centro. Assim, os autores afirmam que, se o SN alvo da relativização for maximamente diferente do SN interveniente, no domínio da OR, até a saturação de seu respectivo predicador, então o processamento dessa estrutura será menos complexo do que o de estruturas que apresentem SNs semelhantes. O modelo de Gibson pode ser considerado mais explícito do que o de Gordon et al. justamente porque fornece uma métrica explícita capaz de medir a diferença entre os SNs envolvidos no processamento de ORs.

### **3. Experimento 1: julgamento de aceitabilidade**

Com o objetivo de testar, com dados do português brasileiro, as previsões do modelo de Gibson (2001), foi conduzido, no âmbito do GEPEX (Grupo de Estudos e Pesquisas em Linguística Teórica e Experimental) da Universidade Federal Fluminense<sup>2</sup>, um experimento *off-line* denominado julgamento imediato de aceitabilidade por meio de

---

<sup>2</sup>Queria visitar o site [www.gepex.org](http://www.gepex.org). Agradecimentos aos bolsistas IC-CNPq Juliana Benevides e Luiz Guimarães, que ajudaram a programar e aplicar os experimentos aqui descritos.

notas em escala. Trata-se de um paradigma experimental clássico na pesquisa em psicolinguística e em sintaxe experimental (cf. COWART, 1997), cuja tarefa consiste em apresentar ao participante, na tela de um computador, uma dada frase e solicitar-lhe que atribua, tão rapidamente quanto possível, uma nota à frase lida, conforme sua percepção de maior ou menor aceitabilidade.

### 3.1. Desenho experimental

Trata-se de um ensaio randômico, com apenas uma variável independente – a natureza do SN no interior da segunda OR – realizada em três níveis, de acordo com a ilustração que se segue.

/
I) Pronome O repórter que o senador que <b>[eu]</b> conheço atacou desagradou o editor.
II) Nome próprio O repórter que o senador que <b>[João]</b> conhece atacou desagradou o editor.
III) SN indefinido O repórter que o senador que <b>[um estudante]</b> conhece atacou desagradou o editor.

**Ilustração 2. Exemplos das três condições experimentais da tarefa.**

A variável de resposta (variável dependente) foi a nota atribuída numa escala de 1 a 5, em que 1 significava muito difícil de entender; 2 significava difícil de entender; 3 significava razoável de entender; 4 significava fácil de entender e 5 significava muito fácil de entender.

### 3.2. Hipóteses e previsões experimentais

Assumindo as hipóteses do modelo de Gibson (2001), o experimento previa que ORs que apresentassem, na posição de sujeito da segunda OR encaixada, elementos da categoria pronome pessoal dêitico receberiam as notas mais altas no julgamento dos participantes. No extremo oposto, previu-se que figurariam as ORs com SN indefinidos, que, de acordo com o modelo assumido, possuiriam o maior custo de processamento e, assim, deveriam receber as piores notas médias. Intermediariamente, figurariam as ORs

com nomes próprios, que seriam menos aceitáveis do que as relativas com pronomes, porém receberiam notas superiores às conferidas a ORs com SNs indefinidos.

### 3.3. Materiais

Compuseram o experimento um total de 36 sentenças, sendo 4 de cada uma das três condições e 24 sentenças distratoras, isto é, estímulos sem OR que também apresentavam maior ou menor aceitabilidade em função da manipulação de algum outro fator gramatical aleatório.

A distribuição dos materiais entre os participantes da tarefa foi a intraparticipantes, no sentido de que todos os participantes julgaram as mesmas 12 frases experimentais, 4 de cada condição. Para que isso fosse possível e atendendo à exigência metodológica de que um mesmo participante não julgasse a versão exata de uma frase em cada uma das condições do experimento, recorreu-se à criação de um conjunto de sentenças para balanceamento no esquema do quadrado latino. Dessa forma, cada condição possuía um conjunto de frases de modo que um mesmo participante nunca fosse exposto a versões exatas de uma mesma sentença, mas, sim, a 4 estímulos por condição compostos, cada qual, por itens lexicais completamente independentes.

### 3.4. Participantes

Participaram do experimento 50 brasileiros, falantes monolíngues do português. Todos alunos do primeiro período de Letras da UFF, com média etária de 19 anos. 36 participantes eram do sexo feminino e os demais do masculino.

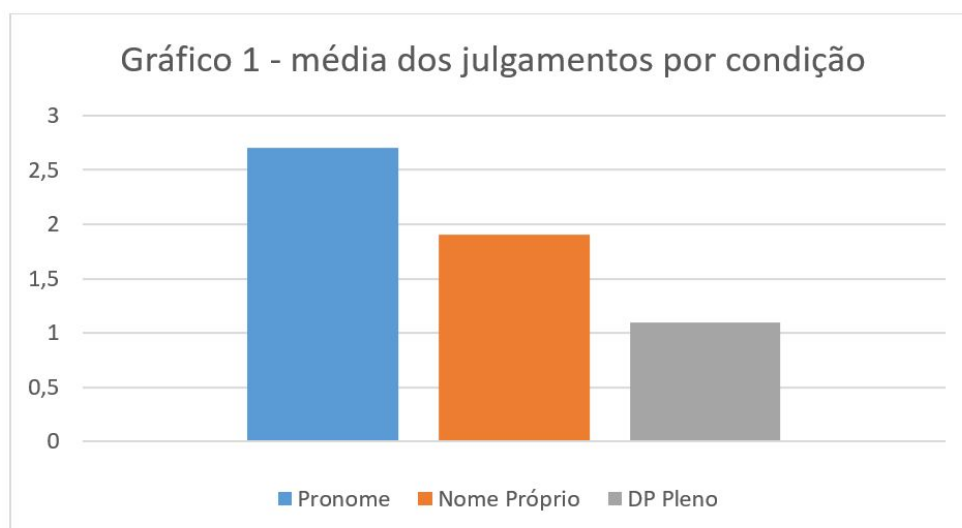
### 3.5. Procedimentos

Cada participante foi conduzido individualmente ao Laboratório do GEPEX, recebeu instruções para a realização do experimento e participou de um treinamento, diante do experimentador, composto apenas por frases distratoras. Após a certificação de que o participante havia compreendido perfeitamente a tarefa e a conseguia executar corretamente no computador, ele passa a ficar sozinho no Laboratório e, então, dava-se início ao experimento.

O participante deveria ler e julgar uma frase por vez. Cada frase permanecia centralizada na tela de um computador durante 5 segundos e, em seguida, desaparecia, dando lugar a três pontos de interrogação. Nesse momento, o participante deveria julgar o mais rapidamente a frase lida, conferindo-lhe, conforme sua percepção, uma das 5 notas descritas em 3.1. Para tanto, o participante tinha em suas mãos uma caixa de resposta com 5 botões, cada um com um adesivo correspondente à cada possibilidade de nota, de 1 a 5. A caixa de resposta também continha uma tecla amarela, que deveria ser acionada pelo participante para que o próximo estímulo fosse apresentado, após a conclusão do julgamento do estímulo anterior. A média de realização do experimento foi de 7 minutos.

### 3.6. Resultados e discussão

SNs indefinidos apresentam nota média 1,1. Nomes próprios atingiram a média 1,9. Pronomes pessoais dêiticos alcançaram as notas mais altas, com média 2,7. Esses dados corroboraram o experimento piloto descrito em Kenedy (2015) e parecem indicar que, apesar de todas as condições terem recebido notas muito baixas, os participantes discriminaram ORs de encaixe recursivo de acordo com a natureza do SN presente na segunda OR.



As médias de cada condição foram consideradas significativamente diferentes entre si no teste qui-quadrado de proporção ( $X^2$  de Pearson), com a estatística  $X^2 = 5,77$ ,  $p < 0,05$ . Tal resultado pode ser interpretado como evidência empírica favorável à atuação da variável “representação do tipo referencial do SN na memória de trabalho” como um fator relevante para explicar, ainda que parcialmente, a dificuldade de processamento –

e, por isso mesmo, a inaceitação – de ORs encaixadas recursivamente ao centro da sentença.

#### **4. Experimento 2: escolha de paráfrase**

Este experimento foi reportado, de maneira informal, em Kenedy (2015). Trata-se do paradigma psicolinguístico *off-line* nomeado “escolha forçada”. Nesse tipo de tarefa, o participante é apresentado a um estímulo e, então, num leque de opções, deve escolher outro estímulo que, de alguma forma, esteja relacionado ao primeiro.

Especificamente, o experimento consistiu em apresentar a crianças do 5º ano escolar ORs de encaixe central único, que variavam conforme a natureza do SN presente, na função de sujeito, no domínio do OR – se pronome dêitico, nome próprio ou SN definido. A tarefa de cada criança consistia em escolher a paráfrase que melhor descrevesse o conteúdo semântico da OR, em termos de valor de verdade, dentre duas opções disponíveis. Numa opção, apresentava-se a paráfrase correta, em que o sujeito da OR era relacionado ao conteúdo do verbo descrito no domínio da OR. Na outra opção, apresentava uma paráfrase incorreta, na qual o sujeito da matriz era apresentado como o agente da ação descrita pelo verbo da OR, que possuía sujeito não correferente ao sujeito da matriz. O objetivo do experimento era verificar se a identificação da paráfrase correta poderia variar de acordo com a natureza referencial do SN sujeito da OR.

##### **4.1. Desenho experimental**

Novamente, trata-se de um experimento caracterizado como ensaio randômico, constituído de somente uma variável independente – a natureza do SN sujeito da OR. Essa variável única possuía três níveis: pronome pessoal dêitico, nome próprio e SN definido. Na ilustração 3, cada uma dessas condições e suas opções de paráfrase são apresentadas.

/
I) Pronome
A menina que você assustou correu.
Paráfrase correta: Você assustou a menina.
Paráfrase incorreta: A menina assustou você.
II) Nome próprio
A menina que Joãozinho assustou correu.
Paráfrase correta: Joãozinho assustou a menina.
Paráfrase incorreta: A menina assustou Joãozinho.
III) SN definido
O aluno que aquela menina assustou correu.
Paráfrase correta: Aquela menina assustou o aluno.
Paráfrase incorreta: O aluno assustou aquela menina.

**Ilustração 3. Exemplos das três condições experimentais da tarefa.**

A variável dependente nesse experimento foi o índice de identificação da paráfrase correta por condição experimental.

#### 4.2. Hipóteses e previsões experimentais

Em acordo com o modelo de Gibson (2001), previu-se que ORs que veiculassem pronomes dêiticos na posição de sujeito apresentariam maiores índices de acertos na identificação da paráfrase correta. Por contraste, as ORs com SN definidos como sujeitos apresentariam o menor número de acertos na identificação correta das respectivas paráfrases. Por fim, esperava-se um nível intermediários de acertos/erros no caso dos nomes próprios, que apresentariam escores maiores do que os SNs definidos, mas inferiores aos dos pronomes.

Embora possa parecer contraintuitivo que crianças do 5º possam ter dificuldades numa tarefa tão simples como a aqui descrita, dados apresentados por, entre outros, Costa et al. (2011) apontaram que infantes precisam de um longo período, até a puberdade, para conseguir processar de maneira completa relações de dependência à longa distância, como as que ocorrem em ORs. Foi essa evidência a motivar o presente design.

#### 4.3. Materiais

O experimento foi composto por 36 sentenças, 4 de cada condição experimental e 24 distratoras. Novamente, dado que a distribuição dos estímulos adotada foi a intra-sujeitos, com medidas repetidas, recorreu-se ao balanceamento no esquema do quadrado latino para que não houvesse repetição de itens lexicais e, assim, que algum participante fosse exposto à exata versão de um estímulo em mais de uma condição experimental. Os estímulos distratores apresentavam diversos tipos de confirmação ou inversão do valor de verdade das sentenças alvo, tais como sinônimos, antônimos e negações.

#### 4.4. Participantes

Participaram do experimento 40 crianças do 5º ano escolar do CIEP Geraldo Reis, o Colégio de Aplicação da UFF, em Niterói. A idade média foi 10 anos e 5 meses. Foi possível distribuir a tarefa para 20 participantes de cada sexo.

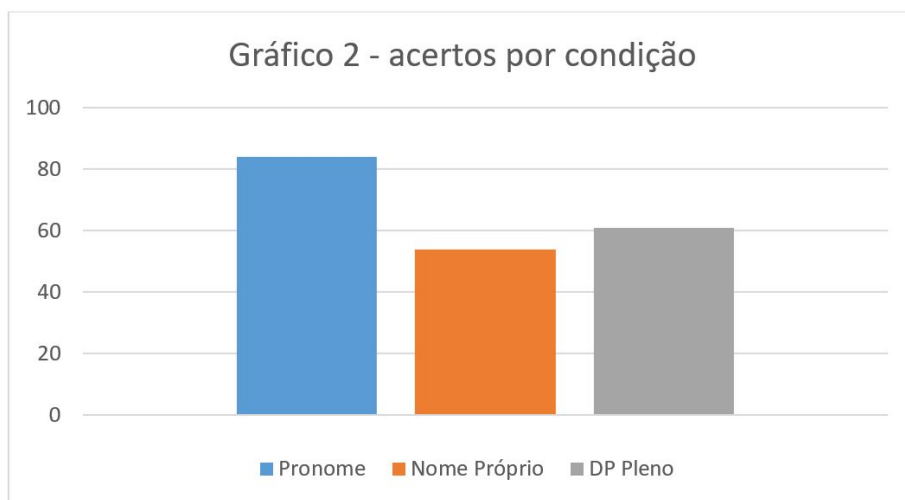
#### 4.5. Procedimentos

Cada participante foi conduzido individualmente a uma sala privada do CIEP Geraldo Reis e recebeu instruções para a realização do experimento, realizando um treinamento composto apenas de estímulos distratores. Após a certificação de que o participante havia compreendido a tarefa, tinha início a aplicação do experimento.

Os participantes recebiam um conjunto e folhas de papel. Cada folha continha a frase alvo e duas propostas de paráfrases, tanto nos estímulos experimentais, como nos distratores. O participante deveria ler, em voz alta, a frase alvo e as propostas de paráfrases para, então, escolher somente uma delas, indicando sua escolha com uma caneta. O tempo médio de realização da tarefa foi 13 minutos. As opções com a paráfrase correta e a incorreta eram ordenadas randomicamente, de modo que a paráfrase correta poderia figurar aleatoriamente como a primeira ou a segunda entre as opções de escolha.

#### 4.6. Resultados e discussão

ORs com pronomes apresentaram 84% de identificação correta de paráfrases, contrastando com somente 54% de acertos na condição SN definido. Nomes próprios registraram 61% de identificações corretas.



Embora SNs definidos (plenos) e nomes próprios apresentem percentuais de acerto diferentes, tal distinção não foi considerada significativa no teste estatístico: ambos foram selecionados pelo participantes de maneira aleatória (nível da chance). Somente a condição pronome pessoal dêitico distinguiu-se das demais na análise do qui-quadrado ( $X^2 = 6,33$ ,  $p < 0.05$ ), indicando uma escolha consistente, para além do nível da aleatoriedade.

Esses resultados, mais uma vez, confirmam as previsões derivadas do modelo de Gibson (2001) e, dessa forma, podem ser interpretados como evidência empírica em favor da hipótese de que ORs de encaixe central (recursivo ou único) tornam-se mais ou menos problemáticas para o processamento psicolinguístico conforme a natureza das expressões nominais presentes no domínio da(s) OR(s). Tais evidências corroboram a interpretação de que fatores relativos à natureza da memória de trabalho humana, com recursos consumidos durante o *parsing*<sup>3</sup>, são responsáveis pela dificuldade de processamento de ORs centrais, possivelmente em associação a outros fatores, como o número de dependências de longa distância não saturadas.

---

<sup>3</sup>*Parsing* é termo técnico da psicolinguística experimental para fazer referência ao processamento cognitivo de sentenças. *Parser* é termo referente ao processador sentencial mental.



## 5. Experimento 3: leitura segmentada automonitorada

O último experimento caracteriza-se como a versão *on-line* das tarefas anteriores. Nele, os participantes tiveram como tarefa a leitura de sentenças que continham orações relativas idênticas às do experimento 2 – portanto, com somente um encaixe central de OR. Os estímulos, nesse tipo de paradigma experimental, são apresentados em segmentos, e não todos de uma vez de maneira global, como no mundo real em textos impressos ou virtuais. Desse modo, torna-se possível aferir o tempo de leitura consumido durante o processamento do segmento que contém a OR, considerando-a isoladamente do restante da sentença. Nesse caso, objetivava-se verificar se o tempo de leitura de uma OR seria variável em função da natureza referencial da expressão nominal presente em seu domínio, na função de sujeito – se pronome pessoal dêitico, nome próprio ou SN pleno definido. O objetivo desse experimento era conferir a possibilidade de ORs com pronomes dêiticos apresentarem latências de leitura menores em relação aos demais tipos de expressões nominais, controlados, naturalmente, fatores outros, como a extensão da relativa em termos de palavras ou sílabas.

### 5.1. Desenho experimental

Mais uma vez, tratou-se de um ensaio randômico, com apenas uma variável independente – a natureza do SN sujeito da OR. Tal variável continha três níveis, projetados nas seguintes condições experimentais: pronome pessoal dêitico, nome próprio e SN definido.

/
<b>I) Pronome</b>
A menina / que você mesmo assustou / correu/ espantada.
<b>II) Nome próprio</b>
A menina / que Mariazinha assustou / correu/ espantada.
<b>III) SN definido</b>
A aluna/ que a menina assustou / correu/ assustada.

Ilustração 4. Exemplos das três condições experimentais da tarefa.

As barras inclinadas indicam o limite entre os quatro segmentos apresentados em cada sentença. A medida dependente (variável de resposta) coletada nesse experimento foi o tempo médio de leitura do segmento que continha a OR – chamado de segmento crítico – que, no caso, foi o segundo segmento. Os tempos de leitura do segmento pós-crítico (terceiro segmento, com o verbo da oração matriz) foi também registrado para verificar possível efeito *spill-over*, que ocorre quando um custo de processamento de um segmento é derramado sobre o segmento imediatamente seguinte e, às vezes, somente lá se torna visível.

## 5.2. Hipóteses e previsões experimentais

Pelo exposto nas seções anteriores, é possível entender que, a partir do modelo aqui adotado, deriva-se a previsão comportamental de que ORs com pronomes pessoais dêiticos demandarão latências de leitura menores, uma vez que impõem menos demanda à memória de trabalho humana consumida durante o *parsing*. Por sua vez, SNs referenciais plenos definidos apresentarão as maiores latências, dado que demandam consumo de mais recursos da memória de longo prazo, com a busca lexical, e também da memória de trabalho, coma representação de entidades ativas no universo do discurso presente. Finalmente, tempos médios entre as condições já citadas são esperados para os nomes próprios, em função dos mesmos argumentos apresentados nos experimentos anteriores.

## 5.3. Materiais

O presente experimento foi composto por 36 sentenças, 12 experimentais, 4 de cada condição, e 24 *fillers*, isto é, duas dúzias de estímulos distratores. A distribuição dos estímulos foi intraparticipantes e, por isso, houve balanceamento de sentenças no quadrado latino.

As ORs continham, todas, o mesmo número médio de caracteres (23), apesar de o número de itens lexical em cada condição não poder ser rigorosamente o mesmo, uma vez que o pronome dêitico é um item monossilábico, com dois ou três caracteres apenas. Por isso, uma palavra a mais foi adicionada à condição pronome pessoal dêitico, de modo

que houvesse um equilíbrio entre as três condições em termos de extensão e material físico para leitura.

#### 5.4. Participantes

Deste experimento participaram 20 brasileiros, monolíngues do português. Todos eram alunos do primeiro período de Letras da UFF, 20 indivíduos de cada sexo, com faixa etária média de 19 anos.

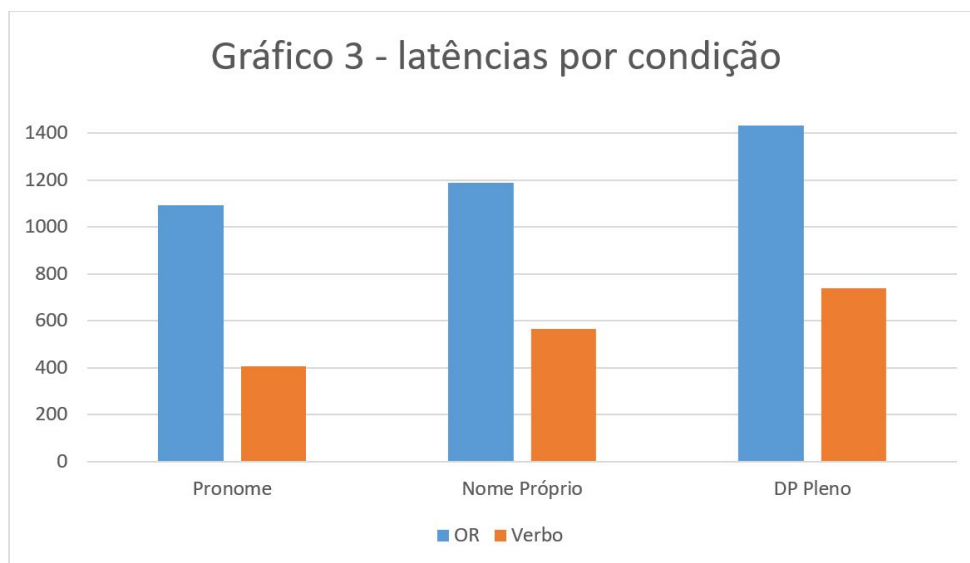
#### 5.5. Procedimentos

Individualmente, cada participante foi conduzido ao Laboratório do GEPEX, onde recebeu instruções para a realização do experimento. Houve um treinamento, diante do experimentador, constituído apenas estímulos distratores. Após a certificação de que o participante havia compreendido perfeitamente a tarefa e a conseguia executar corretamente no computador, ele passa a ficar sozinho no Laboratório e, então, dava início ao experimento.

#### 5.6. Resultados e discussão

Na condição SN definido, foram registradas as maiores latências. Relativas nessa condição foram lidas em médio 1431 milissegundos. Num teste de análise da variância (*one way a Nova*), essa média foi considerada significativamente diferente das latências nas outras duas condições ( $[F_1 (1, 20) = 195,51; p < .05]$ ): nomes próprios atingiram média de 1189 milissegundos, indistintos em termos de significância estatística dos 1092 milissegundos médias da condição pronome dêitico.

Essa tendência de distinção entre a condição “SN pleno” das demais manteve-se, inclusive, quando foram analisadas as latências do segmento pós-crítico – no caso, o verbo da oração matriz. Na condição com SN definido, o verbo da matriz foi lido em médios 738 milissegundos. Interessantemente, quando considerado o segmento pós-crítico, um efeito *spill-over* foi verificado, já que a condição “pronome dêitico” recebeu tempos de leitura significativamente mais rápidos do que as médias da condição “nome próprio”: respectivamente 408 x 567 milissegundos ( $[F_1 (1, 20) = 173,22; p < .05]$ ).



Tais resultados constituem evidência *on-line* em favor das hipóteses de Gibson (2001), considerando-se que o processamento de cada uma das condições experimentais testadas parece variar em função do status referencial do SN que ocupa a função de sujeito de relativa, e não somente em decorrência das estruturas de dependência não local estabelecidas.

## 6. Resultados

Os dados experimentais *off-line* e *on-line* apresentados e discutidos neste texto convergem em favor da hipótese de que o processamento psicolinguístico – e a consequente sensação de (in)aceitabilidade – de ORs encaixadas ao centro, sendo seu encaixe único ou recursivo, é fortemente influenciado por fatores extralinguísticos, notadamente, o custo de manutenção na memória de diferentes entidades do discurso (SNs plenos, nomes próprios, pronomes dêiticos). Tais achados compõem evidência de que a inaceitabilidade universal das ORs de encaixe central recursivo provavelmente decorra das restrições que memória de trabalho humana impõe ao *parser*, e não de fatores de natureza puramente linguística. Isso quer dizer que a rejeição a ORs aninhadas é um fenômeno de performance específico: a representação e a manipulação, na memória, da natureza referencial de SNs envolvidos no processamento de relativas.

Se, por um lado, os dados aqui descritos e interpretados suportam as hipóteses de Gibson (2001) e rejeitam a análise puramente estrutural de Miller & Chomsky (1963), eles não permitem uma distinção entre o modelo de Gibson e o de Gordon et al. (2001).

Com efeito, pronomes dêiticos são diferentes, em termos de morfofonologia e de frequência de ocorrência, de nomes próprios. Nomes próprios, por sua vez, distinguem-se de SN plenos novamente em função de sua massa fônica e de sua frequência. Sendo assim, os modelos de Gibson e de Gordon et al. parecem derivar exatamente as mesmas previsões. Em todo caso, todas as evidências empíricas reunidas em torno do debate sobre ORs de encaixe central recursivo indicam que esse é fenômeno de natureza muito mais psicolinguística do que estritamente linguística.

## Referências

- BENVENISTE, E. *Problemas de lingüística geral*. São Paulo: USP, 1976.
- BEVER, T. The cognitive basis for linguistic structures. IN: HAYES, J. (ed.). *Cognition and the development of language*. New York: Wiley, 1970. pp. 279- 362.
- CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press, 1965.
- COSTA, J. et al. Subject-object asymmetries in the acquisition of Portuguese relative clauses: adults vs. children. *Lingua* 121.6; 2011.
- COWART, W. *Experimental syntax: applying objective methods to sentence judgments*. London: Sage Publications, 1997.
- GIBSON, E. The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In: MARANTZ, A.; MIYASHITA, Y. & O'NEIL, W. (Ed.). *Image, language, brain: papers from the first mind articulation project symposium*, Cambridge, Cambridge: MIT Press, 2000. pp. 95-126
- GORDON, P. et al. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, 2001. pp. 1411–23.
- INGVE, V. A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society*. Vol. 104, n. 5., 1960. pp. 444-466.
- KENEDY, E. Modelos interativos no processamento de frases: a teoria da dependência local. *Prolingua* (João Pessoa), v. 10, 2015. pp. 150-162.
- KIMBALL, J. Seven principles of surface structure parsing in natural language. *Cognition*, Volume 2, Issue 1, 1973. pp 15-47.
- MACDONALD, M. et al. Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101: 1994. pp. 676–703.
- MILLER, G.& CHOMSKY, N. Finitary models of language users. In. LUCE, R. et al. (Ed). *Handbook of mathematical psychology*. New York: Wiley, 1963. pp. 419-461.

**Abstract**

In this paper, we present some experimental results concerning the central and recursive embedded relative clauses. Such relative clauses are intriguing because unlike what happens with right recursively embedded relative clauses, central recursively embedded relative clauses are linguistically unprocessable in any natural language. The reasons for such psycholinguistic impossibility are still under debate (see INGVE, 1960, MILLER & CHOMSKY, 1963, MACDONALD ET AL., 1994; GIBSON, 2001 and GORDON et al., 2001, for a historical perspective). The off-line and on-line experimental data presented and discussed here converge in favor of the hypothesis according to which the psycholinguistic processing - and the consequent sense of (in)acceptability - of central embedded relative clauses is strongly influenced by extralinguistic factors, particularly the memory cost for different discourse entities representation. Such findings compose evidence that the universal unacceptability of central (and recursive) relative clauses is probably due to the constraints that human working memory imposes over the parser rather than purely linguistic factors.

**Keywords:** clauses embedding, relative clauses, recursion, central embedding.