

O processamento do numeral reduplicado *sohoji-sohoji* em Karajá: uma averiguação de ERP durante a compreensão de sentenças¹

The processing of the reduplicated numeral sohoji-sohoji in Karajá: An ERP investigation during sentence comprehension

Cristiane Oliveira da Silva

Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro / Brasil
cristianeolivers@gmail.com

Resumo: Este trabalho tem como objetivo averiguar o processamento do numeral distributivo *sohoji-sohoji*, formado pela reduplicação do numeral *um* em Karajá, sob uma perspectiva transdisciplinar aproximando os campos da linguística teórica e da neurobiologia. O experimento de neurociência da linguagem contribui para elucidar a complexidade do processamento de sentenças com *sohoji-sohoji* quando pareado com diferentes cenários quantitativos. Utiliza-se a técnica de extração de ERP (potenciais bioelétricos relacionados a eventos) para observação e extração dos dados de latência visando a análise estatística dos dados cronométricos monitorados *on-line* por um eletroencefalograma (EEG). Os resultados das análises de variância e teste *t* demonstraram que a leitura evento-distributiva é a *default* porque a violação de exaustividade dos itens dos eventos apresentou maior complexidade de processamento do que a violação de exaustividade de participantes. Pluralidade, cardinalidade e diferenciação de eventos também apresentaram efeitos significativos durante a compreensão de sentenças com *sohoji-sohoji*.

Palavras-chave: Karajá; distributividade; ERP; processamento quantificacional.

¹ Trabalho baseado na tese de doutorado defendida na UFRJ, em 2016, sob a orientação de Marcus Maia e coorientação de Aniela Improta França.

Abstract: The aim of this paper is to investigate the processing of the reduplicated numeral *sohoji-sohoji* formed by the reduplication of the numeral *one* in Karajá, under a transdisciplinary approach, bringing together formal linguistic and neurobiology. The neuroscience of language experiment conducted sheds new light into the quantification processing complexity of *sohoji-sohoji* when different quantitative scenarios were previously presented. We use the ERPs extraction technique (Event-Related Brain Potential) to observe and to extract the wave's amplitude in order to conduct statistical analysis from the chronometric data monitored online by the electroencephalogram (EEG). The results from statistical analysis conducted (ANOVA and *t* test) suggest that the event-distributive interpretation is the first activated because the violation of event items exhaustiveness was more complex than violation of subject exhaustiveness. Plurality, cardinality and event differentiation also displayed significant effects during Karaja distributive sentence comprehension.

Keywords: Karaja; distributivity; ERP; quantification processing.

Recebido em 12 de janeiro de 2017

Aprovado em 08 de maio de 2017

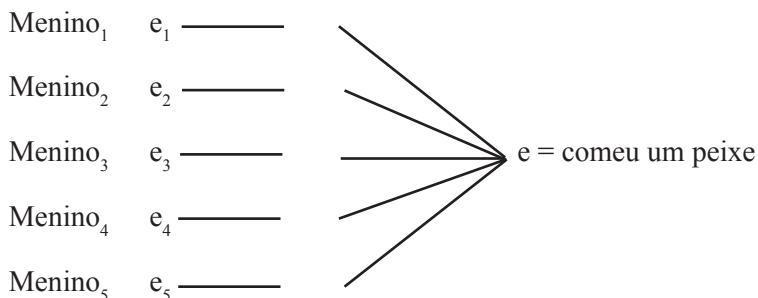
1. Introdução

A quantificação de eventos ou indivíduos é uma operação comum nas línguas naturais. Entretanto, cada sistema linguístico possui mecanismos distintos para expressar noções relacionadas à quantidade. O estudo dos quantificadores das línguas naturais é tema de interesse de diferentes áreas do conhecimento, tendo a filosofia e a lógica, por exemplo, uma gama de trabalhos clássicos que buscam averiguar o funcionamento de elementos linguísticos que exprimem quantidade. Neste trabalho, apresentaremos um estudo de processamento que utilizou a técnica de ERPs (*Event-Related Brain Potentials*) para averiguar as respostas neurobiológicas durante o processamento de sentenças quantificadas com o numeral reduplicado distributivo *sohoji-sohoji* na língua indígena Karajá.

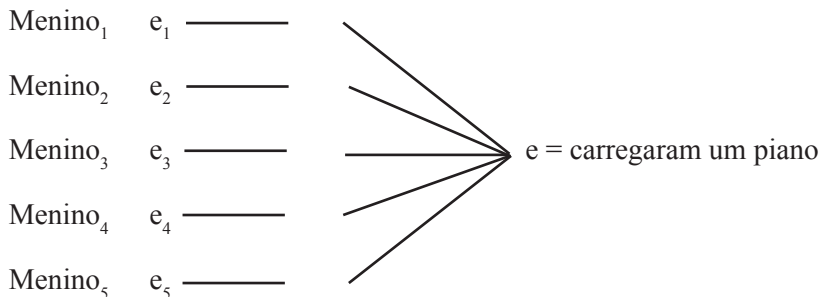
Quantificadores distributivos são responsáveis por uma operação de mapeamento entre os subeventos e os indivíduos de um enunciado.

Uma leitura distributiva prototípica deveria, portanto, operar sobre indivíduos e subeventos, sob a condição de diferenciação entre indivíduos e subeventos mapeados. Como demonstrado no exemplo (1a), para cada indivíduo x tal que x seja *menino* haverá um subevento diferente de *comer um peixe*.

(1a) Cada menino comeu um peixe.



(1b) Os meninos carregaram um piano



Contudo, delimitar determinados tipos de eventos parece não ser uma tarefa trivial e a não diferenciação dos subeventos poderia, então, resultar em uma leitura coletiva, como no exemplo (1b). Tunstall (1998), destaca ainda a existência de uma gradiência entre leituras distributivas e coletivas. Ou seja, haveria uma escala contínua entre leituras distributivas e coletivas prototípicas em que estariam disponíveis as leituras intermediárias, que seriam parcialmente distributivas ou parcialmente coletivas. Essas leituras intermediárias são tratadas por Kratzer (2003, 2007) como cumulativas.

Krifka (1998) e Kratzer (2003, 2007) argumentam a favor da cumulatividade universal dos itens lexicais simples, embora isso não garanta, de fato, que o léxico seja cumulativo na sintaxe da língua. Nessa perspectiva, assumindo que em Karajá itens lexicais sejam cumulativos também na sintaxe, a leitura cumulativa estaria sempre disponível nessa língua. Dessa forma, poderíamos especular que o processamento de uma leitura cumulativa apresentaria menor custo de processamento, exceto quando há um elemento que delimite o escopo quantificacional da sentença, como, por exemplo, um operador distributivo. Logo, ao utilizar um numeral reduplicado *sohoji-sohoji*, forçaríamos uma leitura distributiva para o enunciado, mesmo que outras leituras pudessem ser acomodadas. Neste viés, duas questões podem ser levantadas em relação à língua Karajá. (i) Seria possível processar leituras intermediárias na presença do numeral reduplicado? (ii) O pareamento de leituras intermediárias com sentenças que apresentem reduplicação numeral apresentaria maior custo de processamento? Verificamos em um estudo piloto de descrição de imagens que cenários que continham leituras intermediárias poderiam ser descritos pelos participantes utilizando o numeral *sohoji-sohoji*. Tal fato nos levou a elaborar um experimento que permitisse verificar se haveria diferenças de processamento entre as diferentes leituras quando pareadas com o numeral distributivo nessa língua.

Dessa forma, um experimento de ERPs foi aplicado para averiguar o processamento do numeral reduplicado *sohoji-sohoji* em Karajá. O experimento teve como metodologia um *design* de pareamento de imagens e sentenças (*picture matching design*). Buscamos aferir o curso do processamento das sentenças apresentadas por meio da captura dos tempos de reação e das respostas neurobiológicas dos participantes. Para tal, averiguamos o processamento de sentenças que continham o numeral reduplicado *sohoji-sohoji*, após a exposição de cenários distributivos, coletivos e parcialmente distributivos. Como veremos a seguir, cumulatividade, distributividade e cardinalidade exercem efeitos durante o processamento e interpretação de sentenças com *sohoji-sohoji* em Karajá. Tais evidências parecem dialogar com teorias vigentes acerca do processamento quantificacional e da semântica de numerais distributivos já observados em outras línguas naturais.

2. O numeral reduplicado em Karajá (e outras línguas)

A língua Karajá (família Karajá, tronco Macro-Jê) apresenta reduplicação numeral para expressar distributividade, conforme pode ser conferido no exemplo (2) abaixo. A reduplicação numeral traz à tona dois tipos de leituras distributivas, de acordo com as interpretações apresentadas em (3).

- (2) *weryry sohoji-sohoji bola-di r-e-hu-ra*
 menino um-um bola-posp 3A-VT-arremessar-PST.REC
 Lit: ‘Menino um-um arremessou bola-POSP:instrumento’

- (3) a. ‘Menino arremessou bola de um em um’ evento-distributivo
 b. ‘Cada menino arremessou (uma) bola’ participante-distributivo

Enquanto a leitura evento-distributiva (3a) implica que cada indivíduo tenha arremessado a bola em períodos de tempo distintos, a leitura participante-distributiva (3b) não apresenta essa exigência para a interpretação da sentença. Por sua vez, uma leitura coletiva parece não ser permitida, ou seja, uma leitura em que todos os meninos tenham arremessado ao mesmo tempo uma única bola.

A diferença entre as leituras acima parece ser proveniente da ambiguidade semântica engatilhada pelo numeral reduplicado *sohoji-sohoji*. Um dos objetivos deste trabalho é averiguar se durante o processamento de sentenças distributivas em Karajá haveria preferência por alguma das leituras apresentadas em (3) e se uma leitura coletiva seria imediatamente descartada, dada sua incompatibilidade com a semântica do numeral distributivo.

A possibilidade de uma leitura participante-distributiva exige que a operação de distributividade opere sobre individualidades. A impossibilidade de combinação do numeral reduplicado com nomes massivos em Karajá, que não permitem a individuação de entidades, por se tratar de substância, pode ser confirmada pela agramaticalidade do exemplo (4) abaixo. A sentença (4) demonstra que a substância *bèè* (água) é incompatível com a operação de distributividade exercida pelo operador de *sohoji-sohoji*.

- (4) *Tii bèè sohoji-sohoji-my r-e-wy-reri.
 3P água um-um-POSP 3P-VT-transportar-PRS.CONT
 Ele leva água de uma em uma/ Ele leva cada água

A reduplicação numeral como estratégia de operação distributiva parece ser relativamente comum em diversas línguas. De acordo com o site WALS,² das 251 línguas consideradas 85 apresentam a estratégia de reduplicação numeral para denotar distributividade adnominal. Gil (1982) mostra também que todas as línguas parecem dispor de recursos para expressar distributividade adverbial a partir de algum processo de morfologia numeral. Cabe ressaltar, ainda, que estes estudos abrangem pouquíssimas línguas indígenas. Inclusive, a língua Karajá, objeto de estudo deste trabalho, não faz parte do inventário de línguas observado por estes trabalhos.

Outra língua indígena que apresenta numerais reduplicados para expressar distributividade é o Karitiana (língua indígena da família Arikém). Assim como em Karajá, numerais reduplicados em Karitiana geram diferentes possibilidades de interpretação para uma mesma sentença conforme exemplificado abaixo em (5). Muller (2012) assume, a partir de testes sintáticos, que numerais distributivos em Karitiana são operadores distributivos adverbiais. Para a autora, as duas possibilidades interpretativas são o resultado da possibilidade de diferentes modos de individuação dos subeventos na denotação do predicado verbal.

- (5) Õwã nakakot sypomp.sypomp opokakosypi
 MENINO QUEBROU DOIS.DOIS OVO
 ‘Cada menino quebrou dois ovos’ – participante-distributivo
 ‘Os meninos quebraram dois ovos de cada vez’ – evento-distributivo
 Muller (2012, p. 235)

Entretanto, ao analisar a morfologia numeral em Tinglit (família Na-Dene – Canadá), Cable (2014) propõe que a distribuição numeral sempre permita que cenários participante-distributivo e evento-

² World map of languages structures online, acessado em 27 de outubro de 2014.

distributivo sejam licenciados, semelhante ao que encontramos em Karajá e Karitiana.

Entre outros fatos, eu apresentarei aqui que o local de posição sintática do numeral distributivo não tem efeito sobre o tipo de cenário distributivo que a sentença descreve. Isto é, não importa se o numeral distributivo é adverbial ou adnominal, a sentença poderá ser verdadeira tanto para cenários participante-distributivo quanto para cenários evento-distributivos.³

Cable (2014) então propõe, baseado em um teste de descrição de cenários, duas denotações semânticas, uma distributiva-adverbial, outra distributiva-adnominal para a morfologia numeral em Tinglit. Entretanto, o autor argumenta que ambas as denotações podem gerar tanto leituras distributivas sobre eventos quanto sobre indivíduos. Portanto, conclui-se que a ambiguidade gerada por numerais distributivos é recorrente nas línguas que possuem tal estratégia morfológica para expressar distributividade em seus enunciados. Com base nos dados levantados, a ambiguidade dos numerais distributivos não parece ser fruto de aspectos sintáticos e sim de características semânticas particulares a esses elementos.

Assim, com o objetivo de averiguar as possíveis diferenças de processamento entre as leituras possíveis (e impossíveis) para o numeral reduplicado em Karajá, aplicamos um experimento de EEG apresentado na seção 4 deste trabalho. Os resultados obtidos podem nos auxiliar a compreender os mecanismos cognitivos envolvidos durante o processamento da distribuição numeral. O processamento de sentenças que apresentam elementos quantificadores mostra-se um procedimento bastante complexo, uma vez que exige interface entre diferentes módulos cognitivos, como, por exemplo, o sintático e o quantificacional. É, portanto, de interesse para a neurociência cognitiva distinguir os mecanismos neurais subjacentes ao acesso a diferentes aspectos da

³ “Amongst other facts, I show here that the syntactic attachment site of the distributive numeral has no effect upon the kind of distributive scenario the sentence describes. That is, no matter whether the distributive numeral is adverbial or adnominal, the sentence may be true in either participant-distributive or event-distributive scenarios” (CABLE, 2014, p. 8)

memória semântica durante a compreensão das frases (JIANG *et al.*, 2008).

3. Pressupostos teóricos

Todas as línguas naturais conhecidas até hoje possuem mecanismos para quantificação de enunciados. Como dito na introdução, um dos objetivos deste trabalho é compreender a natureza e o processamento de um dos elementos quantificadores em Karajá, o numeral reduplicado *sohoji-sohoji*, a fim de contribuir para o entendimento de como a quantificação funciona nas línguas naturais e quais são os princípios básicos que regem estes elementos no âmbito da faculdade humana da linguagem. A capacidade de quantificar elementos linguísticos é comumente expressa por meio de operadores chamados quantificadores que, segundo a definição dada por Lyons (1977, p. 455), são modificadores que, ao se combinarem com nomes, produzem expressões cuja referência é determinada em função do tamanho do conjunto de elementos, ou em termos de quantidade da substância que está sendo referida. Entretanto, já se sabe que há também diferentes tipos de quantificadores não nominais, como é o caso, por exemplo, de certos advérbios que expressam quantidade.

A partir de uma perspectiva relacional dos quantificadores, podemos afirmar que quantificadores denotam relações binárias entre conjuntos. Os quantificadores parecem ser incumbidos da tarefa de relacionar dois conjuntos de elementos distintos, no caso, o conjunto de indivíduos e o conjunto de propriedades. Sob o viés da Semântica Formal, Heim (1982) define que um quantificador, ao relacionar dois conjuntos de elementos, forma uma tríade na qual ocupa a posição de operador da estrutura quantificada, conforme apresentado em (6) abaixo.

(6) a. Cada macaco comeu uma banana.

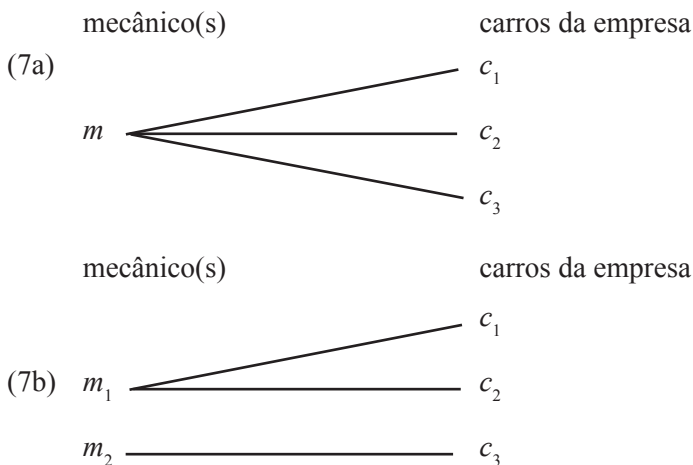
b. para	Cada _x	tal que [x é um macaco]	(\exists_y) tal que y é
	operador	sentença restritiva	escopo nuclear
	uma banana e	x comeu	y

“Para cada x tal que x é um macaco existe um y tal que y é uma banana e x comeu y”.

Portanto, do ponto de vista de análises formais, parece cabível instituir até aqui que quantificadores são operadores que executam a ação de relacionar dois conjuntos pertencentes a classes aparentemente distintas. No caso do exemplo (6), o quantificador *cada* relaciona o envolvimento dos indivíduos, que pertencem ao conjunto dos macacos, em subeventos de “comer uma banana”. Nessa área da linguística, floresceram, principalmente, análises acerca das ambiguidades de escopo causadas, sobretudo, pela interação entre sintagmas nominais quantificados.

(7) Um mecânico consertou cada carro da empresa.

A sentença (7) acima geraria pelo menos duas interpretações possíveis: (7a) há um (único) mecânico que consertou cada carro da empresa; (7b) cada carro da empresa foi consertado por um mecânico (qualquer). Essas duas interpretações podem ser esquematizadas da seguinte forma:



Sintaticamente, cada interpretação acima é representada por uma estrutura de LF (Forma Lógica) diferente que será gerada a partir do lugar de alçamento do sintagma nominal objeto. Logo, para a interpretação de sentenças com NPs quantificados em posição de objeto se faz necessário uma estrutura sintática abstrata que servirá de *input* para a interpretação semântica de sentenças com este tipo de componente estrutural (MAY, 1977, 1985).

A distributividade é uma operação que ocorre, principalmente, quando o núcleo do sintagma determinante é ocupado por um quantificador distributivo e combinado com um predicado. O predicado é, então, entendido como aplicado a cada membro individual do conjunto quantificado. Logo, a operação de um quantificador distributivo é mapear subeventos a cada átomo de indivíduos disponível no discurso. Obviamente, esta é uma definição muito simplória que não dá conta de uma série de mapeamentos distributivos que envolvem, por exemplo, ambiguidade de escopo em sentenças duplamente quantificadas. Entretanto, parece ser uma definição clara e simples o suficiente para explicar a relação entre DPs que contém quantificadores distributivos e eventos aos quais são aplicados, como em “cada macaco comeu banana”. Nesta sentença, “cada” pinça os indivíduos membros do conjunto restritor $\{x:x \text{ é macaco}\}$ sendo o predicado $[[\text{comeu banana}]]$ aplicado a cada indivíduo membro de $\{x:x \text{ é macaco}\}$. Além disso, um quantificador distributivo acaba por pluralizar o evento da sentença em que ocorrem, já que numa estrutura distributiva é exigido ao menos dois subeventos do conjunto de eventos disponíveis, ou seja, exige que a relação R entre indivíduos e eventos seja maior que 1. Logo, para a sentença (6) acima, o significado pode ser interpretado da seguinte forma: existe um evento e e para cada indivíduo m membro de $\{x:x \text{ é macaco}\}$ existe um subevento e' que é parte de e tal que em e' m “comeu banana”.

Quantificadores distributivos como *cada* do português, por exemplo, distribuem sobre entidades individuais. Portanto, ao serem aplicados a sintagmas nominais não contáveis, como aqueles ocupados por nomes massivos do tipo substância, geram sentenças agramaticais, como ilustrado em (8) abaixo, já que não seria possível individualizar suas partes, salvo sob processo de coerção ou quando uma unidade de medida é adjungida ao nome em questão. Em Karajá, o distributivo *sohoji-sohoji* exibe o mesmo padrão de comportamento, já esperado para um quantificador desta natureza, dado a gramaticalidade de (9a) e a agramaticalidade de (9b).

(8) *Ele trouxe cada água.

(9a) *Tii hãdoroó sohoji-sohoji-my r-e-wy-reri.*

3P MURICIUM-UM-PosP 3P-VT-TRANSPORTAR-PRS:CONT

‘Ele leva murici de um em um’/ ‘Ele leva cada murici’

(9b) **Tii bèè sohoji-sohoji-my rewyreri.*

3P ÁGUA UM-UM-POSF 3P-VT-TRANSPORTAR-PRS:CONT

‘Ele leva água de uma em uma’ / ‘Ele leva cada água’

Quantificadores distributivos também forçam uma leitura distributiva do evento (subeventos) da sentença em que está inserido. Portanto, quando combinados a predicados coletivos geram sentenças agramaticais (DOWTY, 1987). Por exemplo, a sentença (10) abaixo, exige que o predicado “estar acumulado no canto da sala” seja aplicado a um conjunto de indivíduos {x:x livros}, com em (10a) e (10b), e não a membros individuais deste conjunto, dada a agramaticalidade de (10c) e (10d).

(10a) Os livros estão acumulados no canto da sala.

(10b) Todos os livros estão acumulados no canto da sala.

(10c) *Um livro está acumulado no canto da sala.

(10d) *Cada livro está acumulado no canto da sala.

E, como esperado em predicados tipicamente distributivos, quantificadores distributivos podem ser aplicados normalmente, como em (11) abaixo. Da mesma forma, caso um quantificador distributivo não seja aplicado a uma sentença cujo predicado seja distributivo, ainda assim um predicado desta natureza forçaria uma leitura do tipo distributivo, mesmo não contendo um DP composto por um núcleo quantificado por um elemento distributivo em sua estrutura, como verificado em (12 e 13) abaixo.

(11) Cada menina dormiu.

(12) Todas as meninas dormiram.

(13) As meninas dormiram.

Assim, de acordo com os dados acima, parece que uma leitura distributiva deve operar não apenas sobre indivíduos > 1, mas também sobre subeventos >1, tendo, portanto, como condição a diferenciação de indivíduos e subeventos – embora seja complicado em certos predicados

delimitar subeventos. A não diferenciação de subeventos resulta em uma leitura coletiva. Entretanto, não existem apenas leituras coletivas e leituras distributivas padrão. De acordo com Tunstall (1998), as estruturas de evento distributivo-coletivas formam uma escala contínua em que leituras parcialmente-coletivas e parcialmente-distributivas estão no centro da escala e as leituras coletivas e completamente distributivas estão nas extremidades. Kratzer (2003, 2007) trata as leituras intermediárias desta escala como cumulativas. Seguindo Krifka (1998), Kratzer (2003, 2007) traz o conceito de cumulatividade como ponto crucial para a denotação de predicados em línguas naturais e argumenta a favor da propriedade de Cumulatividade Universal dos itens lexicais que pode ser conferida no trecho a seguir.

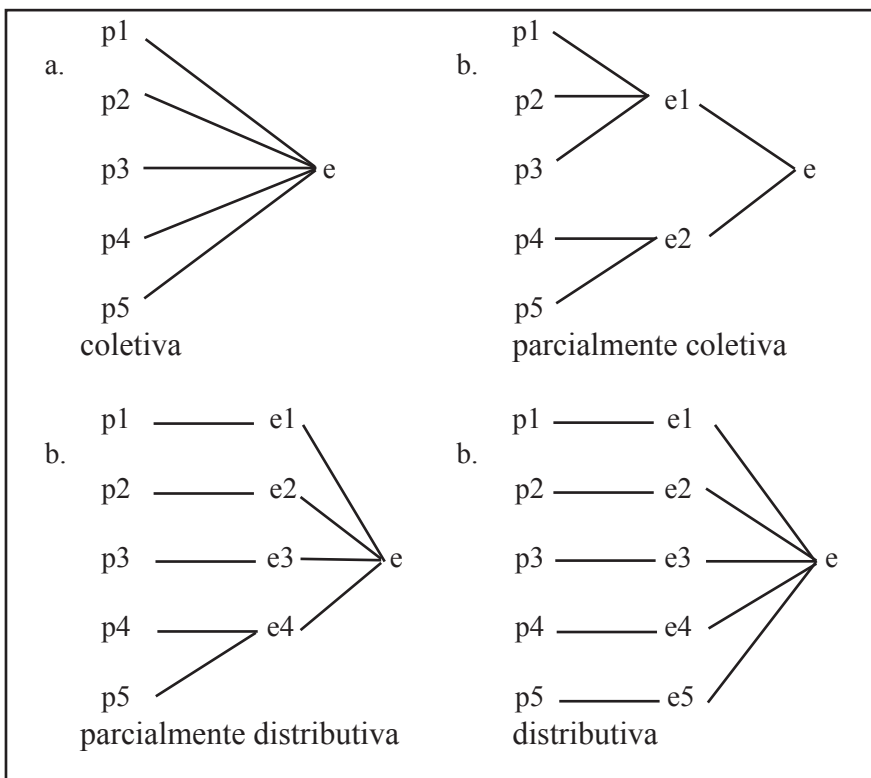
Se um predicado lógico-conceitual básico ‘vermelho’, por exemplo, é verdadeiro para meu chapéu e seu lenço (duas singularidades), então a Cumulatividade Universal diz que esse mesmo predicado também é verdadeiro para a soma de meu chapéu e seu lenço (uma pluralidade).⁴

Ou seja, em predicados simples, a leitura cumulativa estaria, portanto, disponível sem maior custo de processamento, por ser uma propriedade inerente e disponível aos itens lexicais. Entretanto, a presença de um operador distributivo resultaria em uma preferência pela leitura distributiva. Ademais, a cumulatividade inerente aos itens lexicais também explicaria o fato de que uma série de experimentos apresenta evidências de haver maior custo de processamento durante a compreensão de predicados quantificados. Isso ocorreria porque a presença de elementos quantificadores revelaria maior complexidade para o processamento semântico inicial das sentenças, que seriam cumulativas por padrão. O experimento de ERPs apresentado na próxima seção parece apontar nesta direção. Observe abaixo os diagramas abaixo, retirados de Tunstall (1998, p. 97), que buscam escrutinar as leituras intermediárias possíveis entre as leituras coletiva e distributiva padrão. Essa escala serviu para a autora demonstrar que a principal diferença entre a semântica de

⁴ “If there is a basic logical-conceptual predicate ‘red’, for example, that is true of my hat and your scarf (two singularities), then the Cumulativity Universal says that that very same predicate is also true of the sum of my hat and your scarf (a plurality).” (KRATZER, 2003, p. 12)

each e *every* diz respeito à condição de diferenciação, que é mais forte para *each* (todo subevento deverá ser distinto) do que para *every* (ao menos dois eventos deverão ser distintos).

QUADRO 1 – diagramas de leituras intermediárias de Tunstall (1998)



(14) Maria levantou todas as pastas.

Assim, para uma sentença do tipo (14) acima, vamos imaginar o seguinte cenário: há cinco pastas de documento em cima de uma mesa tal que Maria levantou todas essas pastas. Nesse cenário, há uma série de leituras possíveis de acordo com a escala acima exibida. A leitura (a) – **coletiva** – prevê que Maria levantou todas as cinco pastas de uma única vez, ou seja, houve apenas um evento de levantar as pastas. A leitura (b) – **parcialmente coletiva** – sugere que as pastas foram levantadas em pequenos grupos, por exemplo, Maria levantou três pastas e em

seguida levantou mais duas pastas gerando dois subeventos do evento de “levantar pastas”, como explícito no gráfico acima. A leitura (c) – **parcialmente distributiva** – exige que haja subeventos aplicados a membros individuais, mas permite concomitantemente subeventos coletivos, como, por exemplo, Maria levantou três pastas uma de cada vez e, depois, duas pastas ao mesmo tempo, ou seja, quatro subeventos do evento de “levantar pastas”. Por fim, a leitura (d) – **completamente distributiva** – exige que cada pasta seja levantada individualmente gerando, portanto, cinco subeventos de “levantar pasta”.

Na seção a seguir, averiguaremos, entre outras coisas, se cenários que apresentam distributividade parcial podem ser pareados com sentenças que contenham o numeral reduplicado *sohoji-sohoji* e qual seria o custo cognitivo deste procedimento. Além disso, buscamos demonstrar a aceitabilidade (ou não) de diferentes cenários quantificados quando o participante fosse exposto à uma sentença operada por *sohoji-sohoji*. Os resultados *off-line* e *on-line* corroboram para a existência de uma gradiência de aceitabilidade para diferentes leituras distributivas, na qual cenários completamente distributivos e singulares estão nas extremidades opostas da escala. Veremos adiante que, embora os resultados de testes comportamentais tenham elicitado aceitabilidade para os diferentes cenários apresentados, inclusive aqueles tipicamente não distributivos, as medidas de tempo (ms) e amplitude (μ V) parecem fornecer dados significativos para a nossa hipótese de que, embora diferentes leituras possam ser possíveis para o quantificador *sohoji-sohoji*, a leitura distributiva padrão é a primeira ativada por sentenças distributivas em Karajá cujo operador seja *sohoji-sohoji*, ou seja, a que exige menor custo cognitivo durante o processamento da sentença.

4. Processamento do numeral reduplicado *sohoji-sohoji*

Para averiguar as etapas de processamento do numeral reduplicado *sohoji-sohoji*, montamos um design experimental que permitiu a captação de dados *on-line* e *off-line*. A tarefa consistia no pareamento de diferentes cenários controlados, que eram imagens apresentadas na tela do computador, com sentenças distributivas gravadas em áudio. Este *design* é comumente conhecido como pareamento de imagens (*picture-matching design*).

O objetivo deste experimento⁵ foi observar se contextos distributivos padrão apresentariam diferenças de processamento em relação a outros contextos quantitativos (parcialmente distributivo, coletivo, não exaustivo, extra-elemento e unicidade) durante a compreensão de sentenças com o numeral reduplicado *sohoji-sohoji* em Karajá. Isto porque observamos em testagens piloto previamente averiguadas que cenários não prototípicamente distributivos eram aceitos pelos participantes como cenários possíveis para sentenças que continham *sohoji-sohoji*. Portanto, o presente experimento buscou averiguar se a aceitação de diferentes cenários quantitativos para sentenças com *sohoji-sohoji* seria resultado de um processo reflexivo envolvendo variáveis independentes como interpretação e acomodação, posteriores ao parseamento de sentenças, ou se a aceitação de diferentes cenários permitidos pelo quantificador *sohoji-sohoji* estaria relacionada às propriedades semânticas intrínsecas, geradas por este numeral reduplicado e, portanto, passíveis de serem observadas em um nível mais reflexo da compreensão de sentenças, chamado de processamento *on-line*.

4.1 Participantes

Participaram deste experimento 22 sujeitos (16 homens e 6 mulheres), com visão corrigida ou normal, audição normal, destros (OLDFIELD, 1971), falantes nativos da língua Karajá, habitantes das aldeias Btõiry ou Hãwalò (Ilha do Bananal, TO). Todos os participantes declararam não ter ingerido bebida alcoólica ou qualquer outra droga (i)lícita 24 horas antes do experimento e ter concluído o ciclo básico do ensino fundamental. A média de idade entre os participantes foi de 26 anos.

4.2 Metodologia e Materiais

O experimento consistia na apresentação da seguinte sequência de eventos: uma imagem aparecia centralizada em um monitor CRT, seguida por quatro áudios – correspondentes aos seguimentos um, dois, três e quatro de uma sentença em Karajá. Todas as sentenças experimentais

⁵ O experimento foi realizado em parceria com o laboratório ACESIN (Laboratório de Acesso Sintático coordenado por Anieli Improta França), referência na área de neurofisiologia da linguagem no Brasil.

apresentavam a estrutura sujeito-distributivo-locativo-verbo respeitando o padrão tipológico da língua, como no exemplo (15) abaixo. Após ouvir aos áudios, a tarefa dos voluntários era decidir se a imagem seria compatível, ou não, com a sentença ouvida. Este *design* é conhecido na psicolinguística experimental por *picture-matching* ou pareamento de imagens.

(15) *Krukru sohoji-sohoji òra-tyreki r-y-i-reri.*

PÁSSARO UM-UM TRONCO-POSP 3A-VT-RAIZ-PRS.CONT

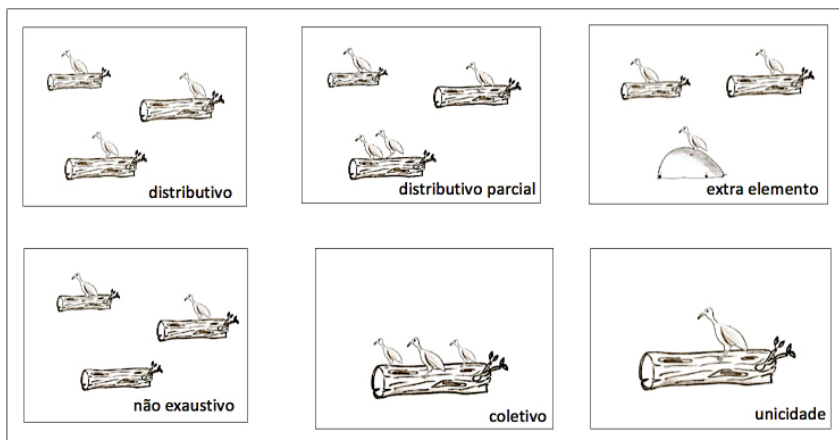
‘Cada pássaro está no tronco’

Para cada sentença experimental, foram produzidos seis cenários, a saber, (i) **distributivo padrão**, (ii) **distributivo parcial**, (iii) **extra-elemento**, (iv) **não exaustivo**, (v) **coletivo** e (vi) **unicidade**. O cenário (i) **distributivo padrão** apresentava a distributividade prototípica. Ou seja, os elementos dos conjuntos de indivíduos e as subeventualidades do conjunto de eventualidades⁶ eram distribuídos em relações de um por um, de forma que o número de relações entre os conjuntos seja maior que 2 e que haja diferenciação entre os subeventualidades. O cenário (ii) **distributivo parcial** demonstrava falha de cardinalidade na operação de relacionar indivíduos e eventualidades, uma vez que a relação de um para um era violada em uma das subeventualidades apresentadas. Os cenários (iii) **extra-elemento** e (iv) **não exaustivo** eram complementares. No primeiro, havia um elemento do grupo de indivíduos que não participava da subeventualidade descrita, já no segundo, havia uma subeventualidade que não estava relacionada a nenhum indivíduo em questão. Esses cenários testaram a propriedade de universalidade/exaustão de indivíduos e eventualidades, respectivamente. O cenário (v) **coletivo** era composto por apenas uma eventualidade para a qual todos os indivíduos eram aplicados, logo testava se propriedade de diferenciação entre subeventualidades era exigida por *sohoji-sohoji* no curso do processamento. Por fim, o cenário (vi) **unicidade** contava apenas

⁶ Em uma análise que emprega tanto predicados estativos como eventivos, o termo eventualidade (introduzido por BACH, 1986) é frequentemente utilizado como uma categorial geral que compreende ambos predicados. ‘In analyses employing both states and events, the term eventuality (introduced by BACH, 1986) is often used for the general category covering both’ (LASERSOHN, 2006).

com um átomo do grupo de indivíduos que, por sua vez, participava de uma única eventualidade, portanto, testava se um contexto singular era imediatamente descartado durante o processamento do numeral distributivo. Por exemplo, no quadro (2) abaixo podemos conferir os seis cenários gerados para a sentença (15). As imagens foram produzidas tendo como referência o banco de imagens produzidos por desenhistas Karajá no âmbito do projeto PRODOCLIN|Karajá.⁷ Cada imagem foi desenhada com lápis HB em folha A4 e digitalizada por meio de equipamento *scanner* em 300-bpi. A partir digitalizações, os cenários foram montados com o auxílio do *software* Keynote para Mac OS e exportados em formato JPG.

QUADRO 2 – Exemplos de cenários para as condições experimentais



Sentença: *Kruku sohoji-sohoji orà-tyreki ryireri*. ‘Cada pássaro está no tronco’

O experimento foi programado no software E-prime (*Psychology Software Tools, Inc. – PST*), plataforma comumente utilizada para apresentação de estímulos linguísticos com acoplamento de EEG monitorado pelo software PyCorder (*Brain Vision LLC*). Foram elaboradas seis versões do experimento em formato de quadrado latino, de

⁷ Projeto de Documentação da Língua e Cultura Karajá, parte do Programa de Documentação de Línguas Indígenas do Museu do Índio (FUNAI), apoiado pela UNESCO.

maneira que nenhum participante viu as mesmas sentenças experimentais, porém todos os participantes viram todas as condições experimentais a partir de diferentes sentenças.

As listas foram apresentadas tanto na ordem normal como inversa, para evitar efeitos de ordem de lista em cada versão. A partir da criação das versões experimentais distribuídas entre sujeitos – *design* de quadrado latino – é possível garantir que um mesmo sujeito nunca veja as mesmas sentenças, mas seja exposto a sentenças de todas as condições. Ao todo, para este experimento, seis listas/versões foram programadas de acordo com a distribuição dos materiais experimentais em formato de quadrado latino. Este protocolo experimental é importante para inibir que fatores extralinguísticos como memória, por exemplo, atuem de modo a interferir no processo de compreensão de sentenças. A distribuição dos materiais entre os participantes também auxilia para que os dados obtidos sejam mais robustos, uma vez que a análise entre sujeitos dilui fatores subjetivos relativos aos indivíduos. Além disso, os estímulos foram pseudo-randomizados dentro de cada lista de cada versão, seguindo as seguintes restrições: (i) estímulos experimentais foram separados por, pelo menos, um item distrator; (ii) nenhuma sentença experimental foi seguida por outra da mesma condição.

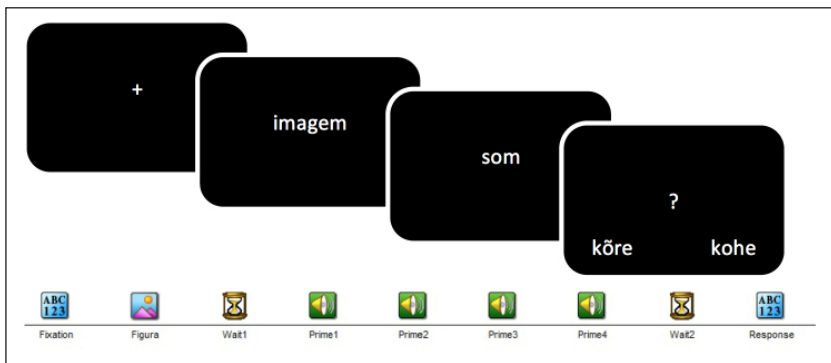
Formam o conjunto total de materiais deste experimento 48 sentenças experimentais acrescidas de 48 distratoras gravadas em áudio e 336 imagens digitalizadas e tratadas. Os áudios foram traduzidos do português para o Karajá em conjunto com dois professores de língua Karajá, conferidos por um terceiro falante, e, finalmente, gravados por um consultora Karajá. Cada sentença experimental era formada por quatro segmentos, respeitando-se o padrão de ordem vocabular da língua Karajá, conforme anteriormente mencionado, e descrito de forma mais detalhada no exemplo (16) abaixo:

(16)	<i>Kruku</i>	<i>sohoji-sohoji</i>	<i>òra-tyreki</i>	<i>r-y-i-reri</i>
	PASSARINHO	UM-UM	TRONCO-EM	COP
	sujeito	distributivo	locativo	verbo
	seg1	seg2	seg3	seg4

Aos participantes, cabia a tarefa de observar a imagem e ouvir a sentença a que seria exposto por meio de monitor acoplado simultaneamente a um computador PC e a um aparelho de eletroencefalografia. Ao fim

de cada par <imagem,som>, a tarefa era avaliar se os materiais eram compatíveis entre si ou não. Para isso, os participantes utilizaram um teclado em que havia duas teclas disponíveis destacadas pelas cores verde e vermelho. Caso o participante julgasse que a sentença ouvida era compatível com a imagem anteriormente vista, apertava a tecla verde. Se julgasse que a sentença era incompatível com a imagem vista, apertava a tecla vermelha.

FIGURA 1 – Sequência de telas apresentadas aos participantes durante experimento de EEG



Conforme ilustrado pela figura (1) acima, a apresentação do conjunto de estímulos começava com uma cruz de fixação no centro da tela por 500ms, seguida pela imagem/cenário experimental por 1500ms. Em seguida, os quatro segmentos em áudio eram apresentados. Os segmentos foram controlados de modo a apresentarem um tamanho médio de 1500ms cada. O tempo total entre os segmentos foi de aproximadamente 6000ms. Após ouvida a sentença, havia uma pausa de 250ms, seguida da tela de resposta, que tinha um *time-out* de 2000ms, ou seja, mesmo que o participante não apertasse nenhum dos botões de resposta, após dois segundos um novo conjunto de dados seria apresentado. Ao final de cada sequência experimental, uma tela em branco de 1000ms era apresentada antes de passar para um novo conjunto de dados <imagem, sentença>.

A fim de responder se havia ou não compatibilidade entre figura e sentença, os voluntários foram orientados a apertarem um dos dois botões disponíveis para resposta. Os botões de resposta foram balanceados entre os participantes (direita/esquerda) a fim de evitar efeito de lateralidade,

ou seja, impedir que aqueles que já associassem o *certo* com o lado *direito* se beneficiassem dessa associação e vice-versa. Os tempos de resposta foram registrados pelo *software E-prime (Psychology Software Tools, Inc.)*.

Os sinais de EEG foram gravados continuamente pelo *software PyCorder (Brain Vision LLC)* a partir de um sistema de 64 eletrodos acoplados em uma touca plástica (*Brain Products GmbH*) em acordo com o sistema 10-20 estendido (GOMES, 2014). Os eletrodos foram alocados em acordo com o padrão internacional (*International System Locations*), incluindo cinco eletrodos ao longo da linha média (FPz, Fz, Cz, Pz e Oz) e dezesseis canais lateral/temporal, oito em cada hemisfério do cérebro (FP1/ FP2, F3/ F4, F7/ F8, C3/ C4, T3/ T4, T7/ T8, P3/ P4, e P7/ P8). Além disso, outros quarenta e três canais do sistema 10-20 estendido foram utilizados (AF3/ AF4, F1/ F2, F5/ F6, FC1/ FC2, FC3/ FC4, FC5/ FC6, FT7/ FT8, C1/ C2, C5/ C6, CP1/ CP2, CP3/ CP4, CP5/ CP6, TP7/ TP8, P1/ P2, P5/ P6, P7/ P8, PO3/ PO4, PO5/ PO6, PO7/ PO8, CB1/ CB2 e O1/ O2).

Os eletrodos foram referenciados *on-line* pelo eletrodo *terra*, localizado na região frontal, e posteriormente re-referenciados pela média dos canais mastoides (TP9/TP10 – esquerdo/direito). As impedâncias foram mantidas abaixo de 10 k Ω . O EEG foi amplificado e digitalizado em uma frequência de 1000Hz. Depois de gravados, a frequência de amostragem dos dados foi mantida em 500Hz e foram utilizados os filtros de passa-baixa de 30Hz e de passa-alta de 0.1Hz., metodologia utilizada pelo laboratório ACESIN em acordo com os padrões internacionais de pesquisas na área (TANNER *et al.*, 2016). Posteriormente, foram geradas médias – *averages* – de ERPs para cada sujeito por eletrodo em épocas abrangendo entre 200ms e 1200ms em relação ao *onset* do estímulo alvo. As épocas caracterizadas por piscadas ou movimento muscular excessivo foram automaticamente rejeitadas pelo *Software Vision Analyzer (Brain Products GmbH)*, programa utilizado para filtragem e promediação dos dados gravados.

A extração de ERPs (potenciais bioelétricos relacionados a eventos) é um instrumento de avaliação cognitiva eletromagnético cuja aplicação se estende às pesquisas linguísticas no campo da neurociência da linguagem desde os anos 1980, tendo sido desenvolvido a partir do eletroencefalograma (EEG) inventado pelo psiquiatra alemão Hans Berger

e utilizado em humanos desde 1929.⁸ A técnica de ERP é considerada muito apurada pois registra com uma precisão de milissegundos respostas do sistema nervoso central à estimulação motora ou sensorial sendo, assim, considerado uma técnica de medição *on-line*. O EEG fornece uma amostragem temporal contínua da atividade elétrica do cérebro. Tal amostragem pode estar relacionada a diferentes tipos de fenômenos linguísticos e é monitorada, com auxílio de software específico para este fim, por um aparelho de EEG acoplado a um computador que expõe os participantes do experimento a uma determinada tarefa linguística pré-estabelecida pelo pesquisador linguista. Após a aquisição dos sinais elétricos produzidos no córtex e monitorados pelo EEG, se faz necessário o tratamento (aplicação de filtros) dos mesmos para que possa ser possível analisar os dados obtidos (GESUALDI; FRANÇA, 2011). A filtragem é, portanto, uma parte essencial e indissociável da pesquisa de ERP.

Os ERPs são compostos por uma sequência de ondas caracterizadas por sua latência, amplitude e polaridade. O ERP geralmente apresenta valor instantâneo de 10 a 1000 vezes menor do que o EEG de fundo e por isso não pode ser visualizado. Para que possa haver a visualização é necessário realizar média de várias épocas.⁹ Este procedimento se justifica por assumir-se o EEG espontâneo como um ruído branco gaussiano de média zero e os ERPs como as únicas respostas que são realmente sincronizadas com o estímulo. Deste modo, o efeito da promediação é aumentar a relação sinal/ruído (SNR), assim permitindo a visualização do efeito específico do estímulo, no caso, linguístico. (MARQUES, 2011, p. 30)

Assim, a promediação é um tratamento matemático dos dados utilizado para consolidar a média de um mesmo trecho, em torno de

⁸ Assim como os métodos utilizados na psicolinguística experimental, a técnica de ERP é igualmente não-invasiva permitindo, dessa forma, sua utilização em diversas pesquisas sobre cognição e linguagem em diferentes grupos humanos, de diferentes faixas etárias, com ou sem patologias e sem prejuízo ou desgaste para o participante do experimento. Para realizar a técnica, basta servir-se dos equipamentos e softwares necessários para a aferição, tratamento e estatísticas dos componentes que se deseje averiguar.

⁹ Épocas ou eventos são janelas de trabalho no contínuo temporal do EEG demarcadas para estudo.

um determinado momento acoplado no tempo, para todos os estímulos averiguados. Esse tratamento tem o intuito de excluir ruídos aleatórios concorrentes ao sinal relacionado ao evento estudado, como, por exemplo, aqueles oriundos de movimentos musculares, piscar de olhos, interferência elétrica causada por algum equipamento na sala de testagem etc. Sendo o ruído aleatório, a média obtida após a promediação tenderia a zero. Portanto, ao final desta operação de “limpeza” emergiria apenas o sinal relacionado ao evento linguístico averiguado. Ou seja, uma grande média (ou média abrangente) das janelas de evento vinculadas ao tempo de exibição dos estímulos elaborados para esta investigação (LAGE, 2005).

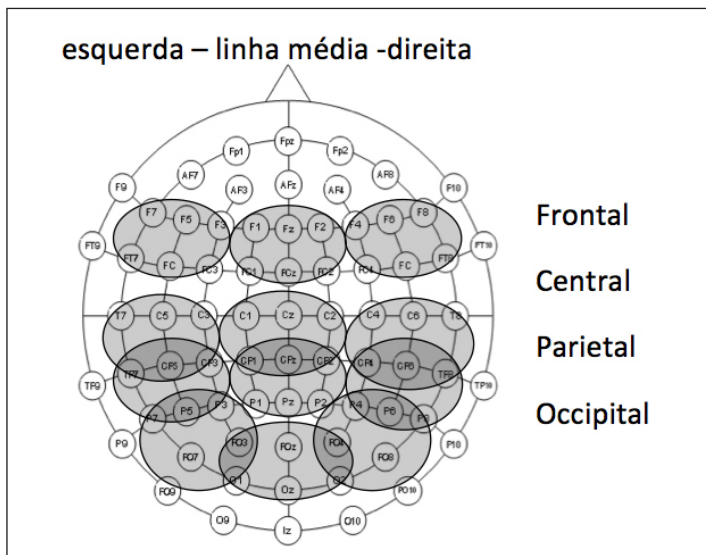
O componente N400 foi descoberto por Kutas e Hillyard no final da década de 70. Os pesquisadores elaboraram um experimento buscando elicitare o componente P3 (já amplamente averiguado na época para reconhecimento de palavras) investigando o papel do contexto no reconhecimento de palavras em sentenças como “*eyebrows*” em “*He shaved off his mustache and eyebrows.*” (Ele barbeou o bigode e a sobrelha dele). Os resultados demonstraram um pequeno componente negativo ao invés de positivo como inicialmente esperado. A fim de observar mais claramente esse efeito, os autores aplicaram um novo experimento com sentenças que continham uma palavra que em vez de ser não esperada pelo contexto era semanticamente anômala. Sentenças como “*He shaved off his mustache and city*” (Ele barbeou o bigode e a cidade dele) e “*I take my coffee with cream and dog*” (Eu tomei meu café com creme e cachorro). Nesse segundo experimento, o efeito de negatividade foi mais explícito uma vez que o mesmo elicitou picos de polaridade negativa bem altos em torno de 400ms no curso do tempo de apresentação da palavra (KUTAS; HILLYARD, 1980). Desde então, este tem sido o componente mais estudado na área da linguagem sendo relacionado principalmente ao conteúdo semântico das palavras per se e às teorias de acesso lexical já que parece fornecer pistas sobre o armazenamento de palavras (FRANÇA *et al.*, 2008). No Brasil, o estudo do componente N400 foi averiguado pioneiramente na tese de doutoramento da professora Anieli Improta França (FRANÇA, 2002) intitulada *Concatenações linguísticas: um estudo de diferentes módulos cognitivos na aquisição e no córtex*. Topograficamente, o N400 tende a apresentar maior ativação na região central e parietal esquerda do cérebro.

Outro componente estudado pela neurociência da linguagem, o P600, foi mais recentemente elicitado por Osterhout e Holcomb (1992) e

desde então tem sido relacionado a anomalias e incongruências sintáticas podendo ser elicitado, por exemplo, em sentenças do tipo **The broker persuaded to sell the stock* (O corretor convenceu a vender o estoque.), em contraste com sentenças bem formadas como *The broker hoped to sell the stock* (O corretor esperava vender o estoque.). Gouvea *et al.* (1995) também encontraram evidências de alta amplitude no componente P600 para sentenças que geram análises temporariamente incorretas (conhecidas na área da psicolinguística como sentenças *garden path*). O P600 também já foi elicitado em anomalias não linguísticas em áreas como música (PATEL *et al.*, 1998) e matemática (MARTÍN-LOECHES *et al.*, 2006). Na área da linguagem, além de violações sentenciais, o componente de P600 parece refletir também violações sintagmáticas como, por exemplo, violações de concordância de número, tempo, gênero e caso, conforme mencionado por Gouvea *et al.* (1995) e investigado por Gomes (2014), Gomes e França (2015) e Gomes e França (2013). A positividade tardia (*Late Positive Component* ou *Late Positive Complex*) já foi elicitada também em estudos sobre quantificação de enunciados (JIANG *et al.*, 2009; KAAAN *et al.*, 2007), o que parece evidenciar que sentenças quantificadas apresentam processamento mais complexo, pois os efeitos de incongruência quantitacional são capturados em momento posterior ao de outros fenômenos da linguagem já averiguados. Kim e Osterhout (2005) avaliam que esse componente pode ser observado em itens cuja incorreção semântica pode levar a uma reanálise sintática.

Para as análises apresentadas a seguir, componentes de interesse N400 e P600 foram inspecionados visualmente e tiveram seus picos registrados automaticamente pelo Vision Analyzer. Para cada um dos canais, foram capturados o pico de amplitude e a voltagem dos componentes de interesse em todos os segmentos e condições. O componente N400 foi inspecionado dentro de uma janela de 350-550ms (capturando ampla negatividade) e o componente P600 foi inspecionado dentro de uma janela de 550-800ms (capturando ampla positividade) após o *onset* do estímulo alvo. Grandes médias – *Grand-averages* – foram formadas pela promediação das médias dos ERPs de interesse de cada participante.

FIGURA 2 – Regiões de interesse baseadas em proximidade anatômica



As medidas dependentes, isto é, a amplitude e latência dos componentes N400 e P600 foram analisadas a partir de análise de variância (ANOVA) e testes paramétricos (Teste-T). As ANOVAS foram aplicadas entre os fatores *Condição* (distributiva, distributiva parcial, coletiva, extra-elemento, não exaustiva e unicidade), *Lateralidade* (hemisfério esquerdo, direito e linha média) e *Anterioridade* (região frontal, central, parietal e occipital). Os testes paramétricos foram realizados para averiguação de diferença significativa entre a condição distributiva padrão e as demais condições experimentais. Portanto, realizamos quatro análises distintas a partir dos dados de ERPs obtidos:

- i. ANOVA das amplitudes dos picos das ondas para o componente N400 por condição, lateralidade e anterioridade. Teste-t contrastando as médias obtidas com as da condição padrão distributiva.
- ii. ANOVA das amplitudes dos picos das ondas para o componente P600 por condição, lateralidade e anterioridade. Teste-t contrastando as médias obtidas com as da condição padrão distributiva.

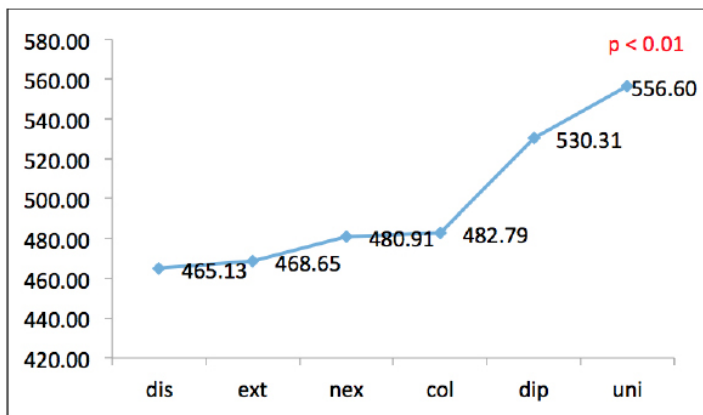
As estatísticas baseadas nos componentes N400 e P600 são as medidas mais discutidas na literatura de neurociência da linguagem

sendo, assim, o melhor ponto de partida para um experimento desta natureza. Além disso, trabalhos que averiguaram diferentes aspectos da quantificação de enunciados elicitaram N400 em sentenças quantificadas, evidenciando um custo de processamento inicial maior para enunciados que envolvem procedimentos como contagem e individuação (CHIARELLI *et al.*, 2011). Também verificamos efeitos de positividade tardia, interpretada por Jiang, Tan e Zhou (2009) como um esforço de integração de elementos quantificadores em sentenças que apresentem inadequação quantificacional. Portanto, para as análises apresentadas a seguir, extraímos a amplitude dos picos de N400 e P600 em todas as condições experimentais a partir das grandes médias dos segmentos 2 e 3 das sentenças manipuladas.

4.3 Dados Comportamentais

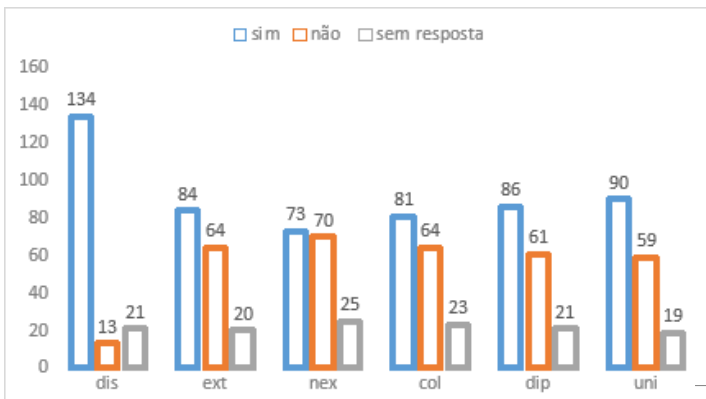
A análise realizada por meio de test-t dos tempos de reação (RT – em milésimos de segundos) obtidos dos participantes nos permitiu observar quais condições apresentaram maior tempo médio de decisão em relação à condição distributiva padrão. Os resultados obtidos apontam que apenas a condição **unicidade** ($t = -2.585$, $df = 21.182$, $p\text{-value} = 0.0172$) elicitou RTs significativamente maiores que a condição distributiva, o que sugere maior custo cognitivo para a decisão dessa condição. Esta análise dos RTs brutos nos mostra que o cenário unicidade foi o que causou maior estranhamento durante a tarefa de pareamento com uma sentença que continha o numeral distributivo *sohoji-sohoji*. As outras condições não apresentaram diferenças significativas, conforme pode ser conferido no gráfico (1) abaixo.

GRÁFICO 1 – RT bruto por condição experimental



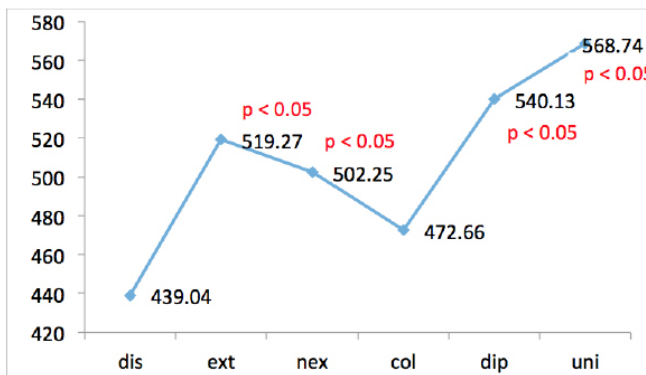
De uma forma geral, observa-se uma tendência em aceitar todas as condições como cenários compatíveis para sentenças com *sohoji-sohoji*, conforme o gráfico 2 abaixo. Entretanto, após a realização de testes de chi-quadrado para cada condição, observou-se que apenas as condições distributiva ($X\text{-squared} = 99.5986$, $df = NA$, $p\text{-value} = 0.0004998$) e unicidade ($X\text{-squared} = 6.4497$, $df = NA$, $p\text{-value} = 0.02249$) apresentaram uma diferença significativa entre as respostas “sim” e “não”. O fato da condição distributiva apresentar o maior índice de aceitação confirma que este é o melhor cenário para as sentenças experimentais averiguadas. A aceitação da condição unidade parece ser resultado de uma tentativa de leitura cardinal do numeral reduplicado, pois como veremos nos dados bioelétricos, essa condição foi a que obteve maior custo cognitivo durante o processamento. Embora as outras condições não tenham capturado um efeito significativo neste teste ($p > 0.05$), observamos que medidas mais reflexas, como os RTs e os potenciais bioelétricos parecem revelar que há diferenças de processamento entre as diferentes condições averiguadas e a condição distributiva padrão. O *time-out* programado para 2000ms gerou alguns itens “sem resposta” que foram excluídos das análises.

GRÁFICO 2 – Índices de resposta por condição



Como em todas as condições obtivemos maiores índices para as respostas “sim”, aplicamos novamente o teste t entre condições para verificar os RTs apenas dos cenários aceitos pelos participantes, excluindo os cenários julgados como não compatíveis. As condições **distributiva parcial** e **unicidade** mantiveram-se com as médias mais altas, neste caso ambas apresentaram diferenças significativas. Entretanto, as condições **não exaustiva** e **extra elemento** também obtiveram tempos significativamente mais altos. A condição coletiva não apresentou diferença de RT significativa quando aceita pelo falante, neste caso a média foi estatisticamente semelhante a da condição distributiva padrão. Os p-valores e médias podem ser conferidos no gráfico 3 abaixo.

GRÁFICO 3 – RT dos pareamentos aceitos por condição experimental



Esses resultados evidenciam a importância dos dados cronométricos para nossa análise, pois os índices de aceitabilidade não mostraram ser um fator decisivo dentre todos os cenários apresentados. Além disso, embora a condição desviante unicidade tenha sido aceita como cenário válido para a sentença experimental, a análise dos RTs evidenciam que essa aceitabilidade foi efetuada em um tempo significativamente maior que para a condição distributiva padrão. Sem a análise dos tempos de decisão em milissegundos também seria impossível capturar a distinção significativa entre as demais condições experimentais observadas, já que a aceitabilidade se manteve constante para todos os cenários. Entretanto, a partir dos RTs ainda não conseguimos obter evidências de que haja diferença de processamento entre as condições não-exaustiva e extra-elemento, nem capturar se há esforço cognitivo maior para aceitar a condição, como esperado.

Os resultados comportamentais apresentados nesta seção foram capturados após o processamento das sentenças, ou seja, de maneira *off-line*. Para análise de dados *on-line*, capturamos também as respostas eletrofisiológicas obtidas através da técnica de extração de ERPs. Esta técnica teve como finalidade observar o processamento das sentenças experimentais de momento a momento durante o curso temporal de compreensão e pareamento das sentenças. Os resultados obtidos a partir dessa metodologia serão apresentados na próxima subseção e apontarão novas reflexões no que concerne a compreensão do processamento do distributivo *sohoji-sohoji* e revelarão dados interessantes sobre como se dá a aceitabilidade dos diferentes cenários averiguados.

4.4 Dados de ERP

Para a análise estatística, utilizamos primeiro o modelo three-way ANOVA (*within subjects*) que incluiu os fatores condição, lateralidade e ROI, com correção dos valores apresentados (*Greenhouse-Geisser*). Após tratamento dos dados elicitados, observamos que o componente N400 apresentou efeito principal significativo para o fator *condição* ($F(5,25) = 7.56 p < 0.0000003$) já no segundo segmento das sentenças experimentais, correspondente ao numeral reduplicado *sohoji-sohoji*. Por outro lado, o componente P600 elicitou efeito principal para o fator *condição* ($F(5,25) = 14.7 p < 0.0017565$) apenas no segmento 3, correspondente ao sintagma posicionado (PP) locativo. Não foram encontrados efeitos principais

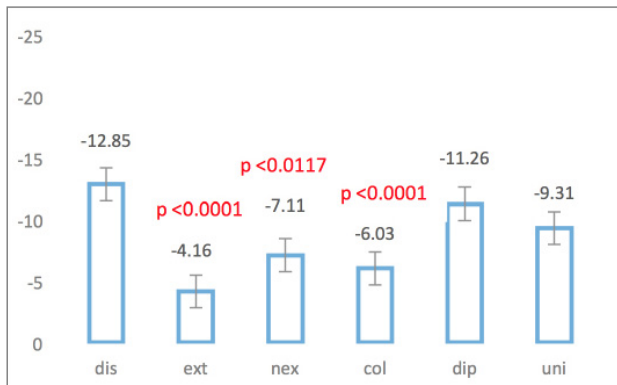
para os fatores ROI ($F(11,55) = 1.19$ $p < 0.3472345$) ou lateralidade ($F(2,10) = 1.90$ $p < 0.2238272$) no componente P600 nem para os fatores ROI ($F(11,55) = 1.41$ $p < 0.195080$ $p < 0.2820406$) ou lateralidade ($F(2,10) = 2.55$ $p < 0.1680510$) no componente N400. Tampouco encontramos interação entre os fatores mencionados. Assim, parece que apenas o fator condição parece influenciar, de fato, as análises entre os resultados obtidos. Por isso, optamos por realizar uma análise incluindo todos as ROIs para esse artigo. Futuramente, outras análises poderão ser realizadas para confirmar as tendências aqui apresentadas.

Posteriormente, foi aplicado um teste t que contrastou a condição distributiva padrão com as outras condições experimentais. Abaixo apresentamos os resultados encontrados para as medidas de amplitude¹⁰ (μV) dos dados de ERPs das Grandes Médias para os componentes N400 e P600.

4.4.1 Componente N400 – negatividade após *onset* de *sohoji-sohoji*

O componente N400, averiguado após o início do segmento que continha o numeral reduplicado *sohoji-sohoji*, apresentou resultados significativos ao comparar a condição distributiva padrão ($-13 \mu V$) com a condição **não exaustiva** ($-7 \mu V$, $t(120)=2.56$ $p < 0.0117$), **coletiva** ($-6 \mu V$, $t(120)=4.88$ $p < 0.0001$) e **extra elemento** ($-4 \mu V$, $t(120)=6.13$ $p < 0.0001$). As condições unicidade e distributiva parcial não apresentaram diferenças significativas.

GRÁFICO 4 – Médias de amplitude do componente N400 por condição

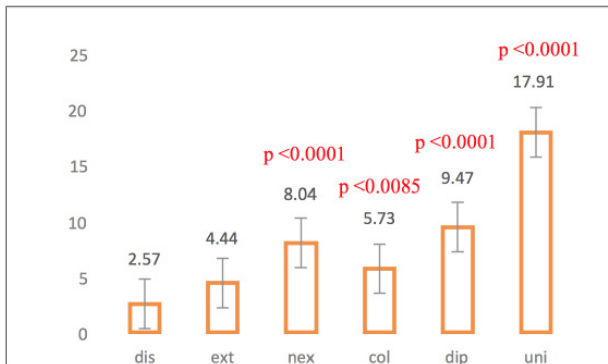


¹⁰ Média dos picos de amplitude dos componentes em microvolts por condição.

4.4.2 Componente P600 – positividade tardia após o *onset* do PP locativo

O componente P600, averiguado após o início do segmento que continha o PP locativo da eventualidade, elicitou diferença significativa entre as condições **unicidade** (18 μ V, $t(118)=17.75$ $p < 0.0001$), **distributiva parcial** (9 μ V, $t(118)=4.91$ $p < 0.0001$), **não exaustiva** (8 μ V, $t(118)=4.17$ $p < 0.0001$) e **coletiva** (6 μ V, $t(118)=2.67$ $p < 0.0085$) ao compararmos com a condição distributiva padrão (3 μ V).

GRÁFICO 5 – Médias de amplitude do componente P600 por condição



4.5 Discussão dos resultados e análises

Este experimento teve como fio condutor a busca por uma compreensão microscópica do processamento do distributivo *sohoji-sohoji*. Para este experimento utilizamos contextos previamente elaborados que foram fornecidos por meio de imagens controladas por *design* específico, como descrito no início desta seção. A metodologia utilizada possibilitou as análises dos dados eletrofisiológicos capturados pelo EEG durante o processamento das sentenças, medida *on-line*, que refletiriam o custo cognitivo durante o processamento das sentenças ao serem pareadas com diferentes cenários quantitativos. Os dados foram então filtrados para a extração e análise das amplitudes (μ V) capturadas a partir da delimitação dos componentes N400 (janela de tempo de 250-500 ms) e P600 (janela de tempo de 550-800 ms). Os resultados obtidos apontam que a aceitabilidade de diferentes condições experimentais em dados comportamentais não reflete o custo cognitivo para o pareamento entre os diferentes cenários apresentados e o numeral reduplicado

sohoji-sohoji, uma vez que os participantes tenderam a aceitar todas as condições. Porém, houve uma tendência de tempos de reação (RTs) mais lentos para cenários desviantes do padrão distributivo, a saber, aqueles que permitiam leituras distributivas intermediárias e não pluralizadas. Portanto, os resultados parecem confirmar a hipótese previamente levantada de que o processamento de leituras cumulativas, apesar de possível, seria mais custoso do que sentenças não cumulativas quando há a presença de um elemento que delimite o escopo quantificacional, neste caso o numeral reduplicado *sohoji-sohoji*. Como discutido abaixo, os dados *on-line* sugerem que as propriedades de cumulatividade, distributividade e cardinalidade exercem efeitos durante o processamento de sentenças com *sohoji-sohoji* em Karajá. Essas evidências parecem dialogar com teorias vigentes acerca do processamento quantificacional e da semântica de numerais distributivos já observados em outras línguas naturais.

No que diz respeito aos dados de amplitude dos ERPs, observou-se que o componente N400 elicitou amplitudes mais baixas para as condições **coletiva** e **extra-elemento** no segmento 2, correspondente ao numeral reduplicado *sohoji-sohoji*. Já o componente P600 elicitou amplitudes mais altas para as condições **não exaustiva**, **coletiva**, **distributiva parcial** e **unicidade** no segmento 3, correspondente ao PP locativo. Esses resultados parecem refletir o custo cognitivo de processamento do numeral reduplicado frente aos diferentes cenários apresentados durante o experimento. O componente N400 revela que o cenário distributivo prototípico, aquele que temos um indivíduo para cada eventualidade, apresenta a maior média de amplitude, seguido pela condição distributiva parcial. Tal evidência parece indicar que o processamento de *sohoji-sohoji* engajaria o sistema cognitivo dos indivíduos em uma operação distributiva, que exige maior custo de processamento em nível reflexo. Já para a medida P600, registrada no segmento locativo que segue *sohoji-sohoji*, observamos que as condições que infringem a leitura do tipo evento-distributivo são aquelas que apresentam maiores amplitudes, revelando que é após o início do procedimento de pareamento do elemento contido na eventualidade que emergem medidas que revelariam quais propriedades seriam mais facilmente processadas em sentenças com numeral distributivo.

A condição **unicidade** foi a que elicitou médias mais altas tanto no componente P600 (18 μ V) quanto nos dados comportamentais brutos

(557 ms) e filtrados por aceitabilidade (568 ms). O que sugere que o numeral reduplicado realmente exige a pluralização do enunciado, no que diz respeito as relações estabelecidas pelos indivíduos e eventualidades disponíveis, em acordo com o que é esperado por elementos distributivos.

A condição **distributiva parcial** também apresentou uma média de amplitude alta para o componente P600 (9 μ V) e também foi a segunda medida mais alta dos tempos de reação filtrados por aceitabilidade (540 ms). Estes resultados indicam que leituras intermediárias, apesar de aceitas, apresentam grande custo de processamento. Neste caso, como *sohoji-sohoji* é formado a partir da reduplicação do numeral *um* também podemos analisar estes resultados como provenientes da falha de cardinalidade numeral apresentada pelo cenário apresentado, ou seja, um problema na operação de contagem.

A condição **não exaustiva** elicitou o componente P600 (8 μ V) e tempos de reação filtrados por aceitabilidade (502 ms) significativos. Nesta condição, um elemento da eventualidade não estava em uma relação R com nenhum indivíduo do cenário. O custo elevado de processamento parece evidenciar que a interpretação de sentenças com *sohoji-sohoji* exige que os subeventos do cenários opere preferencialmente sob o escopo de um quantificador universal.

A condição **coletiva** apesar de elicitou o componente P600 (6 μ V) apresentou médias mais baixas que as outras condições acima mencionadas. Além disso, não foi capturado nenhum efeito nos dados comportamentais. Acreditamos que esse efeito obtido nesta condição poderia ter sido potencializado se fossem averiguados eventos com verbos transitivos que permitissem uma leitura coletiva do evento e não apenas do locativo. Essa hipótese é suportada pelo fato de que nas sentenças averiguadas os participantes aceitaram sem maiores custos cognitivos o cenário coletivo apresentado após o processamento completo da sentença, uma vez que não encontramos efeitos significativos nos RTs.

A condição **extra elemento** não obteve diferenças significativas no componente de P600, ou seja, um cenário em que um indivíduo não participe do evento em distribuição não afetou a amplitude positiva averiguada (4 μ V). Entretanto, nos resultados comportamentais filtrados por acurácia (519 ms) esta condição apresentou resultados significativos. Esses resultados indicam que em um nível mais reflexo, *on-line*, a universalidade sobre eventos é menos custosa do que a universalidade

sobre indivíduos durante a operação de distributividade de *sohoji-sohoji*. A diferença encontrada no RT pode ter sido efeito de uma reinterpretação pós-processamento sentencial.

Em relação aos resultados significativos encontrados no componente N400 para as condições **extra elemento** ($-4 \mu\text{V}$), **não exaustiva** ($-7 \mu\text{V}$) e **coletiva** ($-6 \mu\text{V}$), uma proposta é a de que nestas condições há uma demora maior para o engajamento do sistema de memória em computar a distributividade, pois médias mais baixas refletem menor esforço cognitivo. É, portanto, adequado propor que esses resultados refletem que as incompatibilidades semânticas de universalidade e diferenciação presentes nesses cenários levaria o sistema cognitivo a inicialmente não se engajar na verificação de compatibilidade de tais cenários, procrastinando a verificação para um momento posterior do processamento, conforme verificado pelos resultados obtidos no componente P600 e nos dados comportamentais. A negatividade elicitada pelo componente N400 para as condições **distributiva padrão** ($-13 \mu\text{V}$) e **distributiva parcial** ($-11 \mu\text{V}$) parece ser resultante do efeito de incrementacionalidade dos quantificadores, que seriam interpretados em estágios, assim como acontece com a negação (URBACH; KUTAS, 2010; e URBACH; DELONG; KUTAS, 2015). Logo, a resposta neurofisiológica obtida por medidas de amplitude mais altas para condições distributivas do teste, em um momento inicial do processamento da sentença, refletiria que o participante está engajado no procedimento de uma operação de maior custo cognitivo, no caso, o procedimento da verificação de distributividade numeral. Já a negatividade elicitada para a condição **unicidade** ($-9 \mu\text{V}$) pode ser consequência da implausabilidade semântica em se comparar um cenário não pluralizado com uma sentença distributiva, o que geraria uma resposta neurofisiológica do participante que refletiria o susto em processar uma sentença operada por um numeral distributivo após ser exposto a cenário singular, que deveria ser imediatamente descartado.

5. Considerações finais

Este trabalho procurou demonstrar como a disponibilidade de técnicas cada vez mais sensíveis para averiguar como o estudo do processamento da linguagem pode enriquecer as pesquisas de campo tradicionais e análises formais da linguagem. Nesta perspectiva

transdisciplinar, o experimento de EEG para a extração ERPs teve como objetivo averiguar como os participantes processariam sentenças distributivas com o numeral reduplicado *sohoji-sohoji* quando expostos a cenários que apresentavam leituras intermediárias, não prototipicamente distributivas.

Nossos resultados demonstram que os componentes N400 e P600 foram elicitados quando houve incompatibilidades entre cenário e sentença. A reanálise de condições que violavam a pluralidade, cardinalidade, diferenciação ou exaustividade, foi capturada por uma ampla positividade no componente P600, elicitada a partir dos elementos que ocorriam como sintagmas locativos dos eventos. Em um momento anterior, capturamos um efeito de ampla negatividade no componente N400 em condições que engajavam os participantes em um procedimento quantitativo (distributiva padrão e distributiva parcial) ou que descartavam de imediato a plausibilidade entre o cenário investigado e a sentença distributiva (unicidade).

Esses resultados complementam aqueles encontrados nas medidas cronométricas *off-line* que apontam que *sohoji-sohoji* poderia ou atuar como um quantificador ambíguo, permitindo que mais de uma leitura seja licenciada para a sentença (leituras evento-distributiva ou participante-distributiva) ou licenciaria diferentes cenários de verificação para a mesma leitura, em uma perspectiva próxima àquela proposta por Cable (2014) que propõe uma denotação semântica capaz de dar conta dos diferentes cenários de verificação inerentes aos quantificadores distributivos em Tiglit cuja resolução dependerá, sobretudo, do contexto fornecido.

Além disso, se levarmos em consideração a proposta de Choe (1987), de que o operador distributivo é um quantificador universal que tem um conjunto em restrição e um conjunto em escopo, podemos considerar que o numeral reduplicado *sohoji-sohoji* como *default* colocaria a eventualidade no conjunto em restrição, ou seja, a leitura evento-distributiva seria menos custosa. Isto porque é aos membros do conjunto de eventualidades que o operador universal (\forall) deveria ser aplicado. Ou seja, a não exaustividade dos elementos da eventualidade da sentença seria a leitura que apresentaria maior custo cognitivo porque necessita de reanálise, conforme foi constatado nos dados neurofisiológicos.

A distributividade é um conhecimento que faz parte da cognição humana e é encontrada em diferentes línguas naturais por meio de diversos mecanismos morfológicos e sintáticos. Esperamos que este trabalho possa dar visibilidade para o fenômeno de distributividade numeral encontrado nas línguas ameríndias que podem conter propriedades relevantes para o estudo das áreas pertinentes à linguística e, por conseguinte, para o conhecimento da faculdade da linguagem humana.

Referências

CABLE, S. Distributive numerals and distance distributivity in tlingit (and beyond). *Language*, Washington, Linguistic Society of America, v. 90, n. 4, p. 562-606, 2014.

CHIARELLI, V.; YAGOUBI, R.; MONDINI, S.; BISIACCHI, P.; SEMENZA, C. The syntactic and semantic processing of mass and count nouns: an ERP study. *PLoS ONE*, San Francisco, CA, v. 6, n. 10, p. 1-15, 2011.

DOWTY, D. Collective Predicates, Distributive Predicates, and All. In: ESCOL, 3rd, 1987, Columbus. *Proceedings...* Columbus: Ohio State University, 1987.

FRANÇA, A. I. *Concatenações linguísticas: estudo de diferentes módulos cognitivos na aquisição e no córtex*. 2002. Tese (Doutorado em Linguística) – UFRJ, Rio de Janeiro, 2002.

FRANÇA, A. I.; LEMLE, M.; GESUALDI, A.; CAGY, M.; INFANTOSI, A. F. C. A neurofisiologia do acesso lexical: palavras em português. *Revista Veredas*, v. 2, p. 34-49, UFJF, Juiz de Fora, 2008.

GESUALDI, A. da R.; FRANÇA, A. I. Event-related brain potentials (ERP): an overview. *Revista Linguística*, UFRJ, Rio de Janeiro, v. 7, n. 2, p. 24-42, 2011.

GIL, D. *Distributive numerals*. 1982. Dissertation (Ph.D.) – University of California, Los Angeles, 1982.

GOMES, J. N. *Investigating the dissociation between n400 and p600 effects on the syntax semantics interface: an ERP study*. 2014. Tese (Doutorado) – Universidade Federal do Rio de Janeiro, Faculdade de Letras, Rio de Janeiro, 2014.

GOMES, J.; FRANÇA, A. I. A técnica de ERP: investigando a assimetria sujeito-objeto na interface sintaxe-semântica com EEG. *Letras de Hoje (Online)*, Pontifícia Universidade Católica do Rio Grande do Sul, v. 4, p. 152-163, 2015.

GOMES, J. N.; FRANÇA, A. I. The ample semantic scope of minute language computations: an ERP study of words in Portuguese. *Revista da ANPOLL*, Anpoll, v. 34, p. 309-346, 2013.

GOUVEA, A. C.; PHILLIPS, C.; KAZANINA, N.; POEPPPEL, D. The linguistic processes underlying the P600. *Language and Cognitive Processes*, Taylor & Francis Online, v. 25, p. 149-188, 1995.

HEIM, I. *The semantics of definite and indefinite noun phrases*. 1982. Thesis (PhD) – GLSA, University of Massachusetts, Amherst, 1982.

JIANG, X.; TAN, T.; ZHOU, X. Processing the universal quantifier during sentence comprehension: ERP evidence. *Neuropsychologia*, Elsevier, v. 47, p. 1799-1851, 2009.

KAAN, E.; DALLAS, A. C.; BARKLEY, C. M. Processing bare quantifiers in discourse. *Brain Research*, Elsevier, v. 1146, p. 199-209, 2007.

KIM, A.; OSTERHOUT, L. The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, Elsevier, v. 52, p. 205-225 2005.

KRATZER, A. On the plurality of verbs. In: DÖLLING, J.; HEYDE-ZYBATOW, T. (Ed.). *Event structures in linguistic form and interpretation*. Berlin: Mouton de Gruyter, 2007.

KRATZER, A. The event argument and the semantics of verbs. 2003. Manuscrito. Disponível em: <<http://semanticsarchive.net>>.

KRIFKA, M. The origins of telicity. In: ROTHSTEIN, S. (Ed.). *Events and grammar*. Dordrecht: Kluwer, 1998. p. 197-235.

KUTAS, M.; HILLYARD, S. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, American Association for the Advancement of Science, v. 207, p. 203-205, Jan. 1980.

LAGE, A. C. *Aspectos neurofisiológicos de concatenação e idiomacidade em português do Brasil: um estudo de potenciais bioelétricos relacionados a eventos linguísticos (ERPs)*. 2005. Tese (Doutorado em Linguística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

- LYONS, J. *Semantics*. London: Cambridge University Press, 1977. 2 v.
- MARQUES, F. *Eletrofisiologia da idiomaticidade em indivíduos com Síndrome de Asperger: um estudo de ERPs*. 2011. Dissertação (Mestrado) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.
- MARTÍN-LOECHES, M.; CASADO, P.; GONZALO, R., DE HERAS, L.; FERNÁNDEZ-FRÍAS, C. Brain potentials to mathematical syntax problems. *Psychophysiology*, Wiley Online Library, v. 43, p. 579-591, 2006.
- MAY, R. *The grammar of quantification*. 1977. Dissertation (Ph.D.) – MIT, 1977.
- MAY, R. *Logical form: its structure and derivation*. Cambridge, Massachusetts: MIT Press, 1985.
- MULLER, A. Distributividade: o caso dos numerais reduplicados em Karitiana. *Cadernos de Estudos Linguísticos*, Campinas, v. 54, n. 2, p. 223-243, 2012.
- OLDFIELD, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, Elsevier, v. 9, p. 97-113, 1971.
- OSTERHOUT, L.; HOLCOMB, P. J. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, Elsevier, v. 31, n. 6, p. 785-806, 1992.
- PATEL, A. D.; GIBSON, E.; RATNER, J.; BESSON, M.; HOLCOMB, P. J. Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, MIT Press, v. 10, n. 6, p. 717-733, 1998.
- SILVA, C. O. da. *O processamento do numeral distributivo sohoji-sohoji na língua karajá*. 2016. Tese (Doutorado em Linguística) – Faculdade de Letras da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.
- TANNER, D.; NORTON, J. J.; MORGAN-SHORT, K.; LUCK, S. J. On high-pass filter artifacts(they're real) and baseline correction (it's a good idea) in ERP/ERMF analysis. *Journal of Neuroscience Methods*, Elsevier, v. 15, n. 266, p. 166-170, Jun. 2016.
- TUNSTALL, S. *The interpretation of quantifiers: semantic and processing*. 1998. Dissertation (PhD) – University of Massachusetts, Amherst, 1998.

URBACH, T. P.; DELONG, K.; KUTAS, M. Quantifiers are incrementally interpreted in context, more than less. *Journal of Memory and Language*, Elsevier, v. 83, p. 79-96, 2015.

URBACH, T. P.; KUTAS, M. Quantifiers more or less quantify on-line: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, Elsevier, v. 63, p. 158-179, 2010.