

Revista **Linguística**

Volume 16, número 1 de jan. - mai. de 2020

**Organizadores da edição:
Ana Regina Calindro e Marije Soto**

O número Os Estudos Empíricos na Pesquisa Linguística apresenta 11 artigos originais, uma entrevista e uma resenha que refletem a complexidade e diversidade de pesquisas empíricas na linguística brasileira contemporânea.

ISSN: 2238-975X

UFRJ

COMISSÃO EDITORIAL

Editor-chefe

Aleria Lage, Universidade Federal do Rio de Janeiro, Brasil

Editores

Aniela Improta França, Universidade Federal do Rio de Janeiro, Brasil

Andrew Nevins, Universidade Federal do Rio de Janeiro, Brasil

Marcus Maia, Universidade Federal do Rio de Janeiro, Brasil

Editores Honorários

Miriam Lemle, UFRJ

Maria Luiza Braga, Universidade Federal do Rio de Janeiro, Brasil

Editor Fundador

Lilian Ferrari, Universidade Federal do Rio de Janeiro, Brasil

Conselho Editorial

Anthony Naro, Universidade Federal do Rio de Janeiro, Brasil

Fernanda Ferreira Spoladore, Michigan State University, EUA

Gabriela Matos, Universidade de Lisboa, Portugal

Kees Hegenveld, Universidade de Amsterdam, Holanda

Leticia Sicuro Correa, Departamento de Letras - PUC/RJ

Leo Wetzels, Universidade Livre de Amsterdam, Holanda

Luiz Amaral, University of Massachusetts, USA

Maria Armanda Costa, Universidade de Lisboa, Portugal

Maria Cecília Mollica, Universidade Federal do Rio de Janeiro, Brasil

Roberto Almeida, Concordia University, Canada

Ruth Lopes

Organização da Edição

Ana Regina Calindro e Marije Soto

Redação e Assinaturas

Programa de Pós-Graduação em Linguística - Faculdade de Letras da UFRJ

Av. Horácio de Macedo, 2151 - Sala F. 321. Ilha do Fundão Cidade Universitária

CEP 21941-917 - Rio de Janeiro - RJ. E- mail: ppglinguistica@letras.ufrj.br

Editor Operacional

Elir Ferrari, Editorarte / UERJ

Design e Diagramação

Rafael Laplace, Agoobook/Agoodigital

<http://www.agoodigital.com>

Revista **Linguística**

Rio de Janeiro | Volume 16 | Número 1 | jan. - mai. 2020

Os estudos empíricos na pesquisa linguística

Empirical studies in Linguistic Research

UFRJ

Programa de Pós-Graduação em Linguística
Faculdade de Letras UFRJ

Sumário

Apresentação

6 Os estudos empíricos na pesquisa linguística

Ana Regina Calindro e Marije Soto

Entrevista

14 The new neurocognitive science of language: innovative experimental design for studying syntactic and semantic processing in communicative settings, an interview with Katrien Segaerts

Ana Regina Calindro e Marije Soto

Artigos

43 Sobre a constituição de corpora para línguas com poucos recursos (less-resourced languages)

Lilian Teixeira de Souza

62 Avaliando a vitalidade linguística em contextos de multilinguismo: etnografias versus modelos computacionais

Evani Viotti

85 Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência

Bruno Felipe Marques Pinheiro e Raquel Meister Ko. Freitag

108 O papel da morfologia construcional na formação de antropônimos neológicos

Letícia Santos Rodrigues

124 Estudo sociofonético do português falado na Amazônia: uma comparação interdialetoal entre Belém e Macapá

Rosinele Lemos e Lemos e Camila Roberta dos Santos Brito

146 Testing children's knowledge of generic null pronouns

Karina Bertolino

- 168** Estruturas de posse inalienável adjetivada: apontamentos iniciais sobre a sua aquisição em português brasileiro
Fernanda Mendes
- 200** Interferência de verbos de concordância da Libras (II) no uso de preposições na interlíngua de surdos aprendizes de português L2 (escrito)
Aline Mesquita e Heloisa Salles
- 233** O impacto do uso linguístico oral em L2 no reconhecimento de palavras por bilíngues tardios do português e do inglês
Ricardo Augusto de Souza e Jamila Viegas Rodrigues
- 264** Variáveis lexicais e ortográficas no acesso lexical das palavras do português brasileiro
Gustavo Lopez Estivalet
- 278** O processamento de múltiplas dependências não adjacentes em orações relativas
Késsia Henrique e Cristina Name
- Resenha**
- 303** O laboratório na escola e o desenvolvimento da metacognição linguística: uma experiência translacional com leitura
Erica dos Santos Rodrigues

OS ESTUDOS EMPÍRICOS NA PESQUISA LINGUÍSTICA

*Ana Regina Calindro*¹

*Marije Soto*²

Desde os tempos mais remotos, os estudiosos analisavam obras clássicas com o intuito de verificar como o pensamento humano era desenvolvido (cf. ALTMAN, 2008). Logo, a análise de dados, através de textos consagrados, data de muito antes dos estudos empíricos empreendidos pelos fundadores do pensamento linguístico moderno, entre eles, Saussure (1916), Sapir (1921) e Bloomfield (1933). Desde então, os estudos linguísticos baseados em dados se diversificaram amplamente.

Com o tempo, os campos de investigação foram se ampliando e, portanto, demandando um maior grau de interdisciplinaridade. A mescla de saberes aumentou a complexidade e diversidade das perguntas de pesquisa. Ademais, o avanço da tecnologia também permitiu que *novas* perguntas pudessem ser feitas. Deste modo, metodologias centenárias convivem hoje com o que há de mais moderno, apresentando um leque rico e diverso de pesquisas com base em *corpus* histórico, coletas etnográficas, construção e consulta de *corpora* constituídos digitalmente, experimentos psicolinguísticos, análises acústicas por softwares especializados, estudos com neuroimagem, entre outros. Toda essa grande gama de pesquisas linguísticas tem sua vez nesta edição da Revista Linguística. O objetivo deste número, portanto, foi o de reunir uma amostra de pesquisas recentes baseadas nas mais diversas vertentes teóricas de análise e coleta de dados.

Diversas áreas da Linguística usam dados como base ou verificação de hipóteses, tais como: linguística histórica, linguística experimental (psicolinguística e neurolinguística), aquisição da linguagem, sociolinguística, linguística cognitiva, linguística de corpus, entre outras.

1 Profª. Adjunta do Departamento de Linguística e Filologia, Faculdade de Letras, Universidade Federal do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: anacalindro@letras.ufrj.br.

2 Profª. Adjunta do Departamento de Estudos da Linguagem, Instituto de Letras, Universidade do Estado do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: marijesoto@hotmail.com.

Este volume se inicia com dois textos que discorrem sobre a importância da escolha e elaboração de *corpora* com o intuito de formar uma base de dados que sirva para pesquisas futuras. Lílian Teixeira de Souza, em “Sobre a constituição de *corpora* para línguas com poucos recursos (*less resourced languages*)”, traz uma reflexão sobre os critérios estabelecidos pela Linguística de *Corpus* na composição de *corpora* para análises linguísticas em geral. Como, por exemplo, a constituição de documentação de línguas ameaçadas. Com este foco, a autora apresenta considerações sobre os *corpora* eletrônicos disponíveis de línguas com maior presença digital para, então, abordar a importância e relevância em se constituírem *corpora* de línguas com menor presença digital e menos recursos, como línguas indígenas brasileiras.

Em seguida, Evani Viotti, em seu artigo “Avaliando a vitalidade linguística em contextos de multilinguismo: etnografias versus modelos computacionais”, aborda a importância em se considerar questões socioeconômicas e políticas que caracterizam certas ecologias multilíngues ao estudar línguas como da região do Alto do Rio Negro na Amazônia. Neste contexto, a autora traz um panorama interessante sobre as relações sociais dos povos que habitam essa região. Tais considerações são a base para que se defenda a língua como um fenômeno social, parte de um sistema complexo, aberto, dinâmico e auto-organizado. Portanto, segundo a autora, a teoria da complexidade (BECKNER *et al.*, 2009; ELLIS; LARSEN-FREEMAN, 2009; KRETZSCHMAR, 2015; MUFWENE, 2013; MUFWENE; COUPÉ; PELLEGRINO, 2017; entre muitos outros) seria o caminho mais adequado para dar conta das características dinâmicas das línguas.

Seguindo essa vertente no âmbito da valorização e necessidade de ampliar o conhecimento de dialetos de regiões menos estudadas no Brasil - devido a questões sociais, históricas e econômicas (que vão além do escopo deste texto) - apresentamos dois artigos que trazem dados das regiões norte e nordeste do país.

Em primeiro lugar, as autoras Rosilene Lemos e Camila dos Santos Brito, integrantes do projeto AMPER-Amazônia, vinculado ao AMPER-POR, analisaram a variação geoprosódica do português falado na Amazônia Oriental no artigo “Estudo sociofonético do português falado na Amazônia: uma comparação interdialetoal entre Belém e Macapá”. Com base na metodologia do projeto e em outros estudos que abarcam diferentes regiões do Brasil (cf. MORAES, 1993; CRUZ; SEARA; MOUTINHO, 2015), as autoras analisaram o contorno prosódico nos vocábulos oxítonos presentes em sentenças declarativas neutras e interrogativas totais, a fim de observar a variação entoacional da frequência fundamental (F0) em posição nuclear dos enunciados. Os resultados desta pesquisa revelaram semelhanças prosódicas nos falares de Belém e Macapá.

O artigo “O papel da morfologia construcional na formação de antropônimos neológicos” de Letícia Santos Rodrigues, analisa dados históricos da Bahia, mostrando a relevância e importância de pesquisas que levam em consideração a ampla variação linguística existente no Brasil. Com base na Morfologia Construcional (cf. BOOIJ, 2010), a autora analisou antropônimos registrados em fichas de indivíduos associados a uma ordem religiosa de Salvador/Bahia nos séculos XIX e XX. A partir de antropônimos que apresentavam formativos germânicos (como *ald-* ~ *-ald/nald-* ~ *-nald/vald-* ~ *-vald*) foi confirmada a presença do modelo bitemático nos antropônimos neológicos. Além disso, verificou-se que quando os indivíduos eram expostos a um conjunto de antropônimos que se utilizam de dado formativo ou a um antropônimo muito produtivo, são capazes, por analogia, de depreender o modelo e reproduzi-lo ao criar novos antropônimos, seguindo esquemas construcionais pressupostos pela Linguística Cognitiva (cf. FERRARI, 2014).

Ainda no âmbito social e das análises linguísticas baseadas no uso, o artigo “Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência”, de Bruno Marques Pinheiro e Raquel Freitag, apresenta um estudo experimental. Os autores pediram o julgamento da masculinidade e da feminilidade de nomes comuns de dois gêneros de profissões no português brasileiro (como *o/a dentista*), com intuito de verificar se os falantes constroem representações mentais a partir dos estereótipos de gênero. Os autores hipotetizam que essas representações sofrem efeitos da frequência e de saliência - parâmetros que também afetam a gramática. Após a análise dos resultados, verificou-se que a frequência prototípica da profissão interferiu positivamente no julgamento, logo os parâmetros da estereotipia de gênero, medidos pela saliência e frequência, interferem no julgamento das profissões quanto ao gênero nos nomes comuns de dois gêneros. De acordo com os autores, o resultado reflete que o cérebro categoriza informações a partir das experiências linguísticas e sociais dos falantes, em que aspectos probabilísticos são fatores determinantes (BOD; HAY; JANNEDY, 2003).

Outra área da linguística com amplo uso de uma base empírica para as suas constatações é a psicolinguística. Este campo de investigação com foco nos aspectos mentais que possibilitam a aquisição, compreensão e produção de linguagem incorporou as metodologias da psicologia experimental. Nesta revista, há artigos sobre a aquisição de L1 e sobre a aprendizagem de L2, utilizando uma variedade de metodologias psicolinguísticas para verificação de hipóteses.

No artigo “Testing Children’s knowledge of generic null pronouns”, a autora Karina Bertolino aplica uma tarefa de julgamento de valor de verdade a um grupo de crianças brasileiras de quatro anos para observar se já fariam uma leitura genérica de estruturas impessoais, como em *Nessa escola não pode escovar os dentes depois de comer*. A partir de um entendimento do processo de aquisição desse

fenômeno dentro da teoria de Princípios e Parâmetros (HOLMBERG, 2016), ela mostra e explica que crianças de quatro anos já demonstram interpretações genéricas, o que distingue, por exemplo, o português europeu do português brasileiro. Ela ainda detalha um novo caminho metodológico para fazer estudos com crianças a partir dos dois anos, usando um paradigma de olhar preferencial, permitindo, assim, estudos futuros para verificar a interpretação impessoal ainda mais precocemente, pautado em evidências de produção de impessoais nulos de crianças adquirindo estoniano (TORN-LEESIK; VIJA, 2012).

Por um semelhante viés teórico, Fernanda Mendes discute a aquisição de estruturas de posse inalienável adjetivada em português brasileiro, tais como *O João tem os dedos quebrados* e *O João levantou os dedos quebrados*. A autora, que é especialista no assunto, apresenta os resultados de uma Tarefa de Seleção de Figura aplicada a duas faixas etárias (de quatro a cinco anos; de seis a oito), e a um grupo de controle com participantes adultos no artigo intitulado “Estruturas de posse inalienável adjetivada: apontamentos iniciais sobre a sua aquisição em português brasileiro”. Na tarefa, os participantes têm que escolher a figura compatível com sentenças como *Me mostra quem tem a juba laranja*. O objeto da discussão é a interpretação de posse inalienável, disponível ou não de acordo com a presença e/ou tipo de adjetivo que modifica o nome possuído. O motivo para essa questão são restrições dessa natureza em francês (KAYNE, 1975). Os dados, portanto, pretendem investigar se crianças estendem a interpretação inalienável, ou já mostram as restrições características da gramática alvo, como evidenciado pelo comportamento dos adultos.

O estudo com título “Interferência de verbos de concordância da Libras (L1) no uso de preposições na interlíngua de surdos aprendizes de português L2 (escrito)” das autoras Aline Mesquita e Heloisa Salles também parte de uma análise formal. As autoras lançam mão do conceito de interlíngua para investigar a influência de verbos que marcam a sua relação temática com argumentos com movimento, os chamados verbos de concordância em libras, comparando-os com dados sobre a aprendizagem do uso da preposição *para* no português brasileiro na modalidade escrita. Uma tarefa de produção controlada investiga as interferências positiva e negativa na interlíngua de um grupo de surdos adultos em escolas que promovem o bilinguismo (com a modalidade sinalizada para libras, e a escrita para PB).

Continuando a temática do bilinguismo, mas agora para aprendizes de inglês, os autores Ricardo Augusto de Souza e Jamila Viegas Rodrigues mostram como a influência de uso linguístico oral em L2 (inglês) afeta a capacidade discriminatória em pares mínimos (ex.: *them* /ðem/ ‘eles, elas’; *then* /ðen/ ‘então’) por falantes não nativos de inglês em um experimento de rastreamento ocular

nos moldes do paradigma mundo-visual (TANENHAUS *et al.*, 1995). A fundamentação teórica no seu artigo intitulado “O impacto do uso linguístico oral em L2 no reconhecimento de palavras por bilíngues tardios do português e do inglês” se baseia na Teoria de Exemplares (BYBEE, 2001), que explica a importância do uso na construção de representações sonoras do léxico em L2. Dessa forma, os autores não apenas apontam para a relevância de controlar a variável de uso linguístico mediante questionários de auto-declaração em pesquisas de bilinguismos, como também a relevância desse uso para o processo didático.

A psicolinguística e as suas incursões na neurociência também são centrais na resenha e na entrevista. Erica dos Santos Rodrigues assina a resenha do livro *Psicolinguística e Metacognição na Escola* organizado por Marcus Maia. Ela dá o título “O laboratório na escola e o desenvolvimento da metacognição linguística: uma experiência translacional com leitura”, o que remete à aplicação de modelos e metodologias da psicolinguística para diagnosticar e investigar estratégias cognitivas e metacognitivas dos alunos em relação à leitura, trazendo o laboratório de LER - Laboratório de Eletrofisiologia e Rastreamento Ocular (<http://www.ler.letras.ufrj.br/>) literalmente para a sala da aula das turmas de 8o e 9o ano de uma escola estadual. Em grande detalhe, a autora descreve os oito capítulos que apresentam os projetos de pesquisa de uma equipe multidisciplinar de três instituições do Rio de Janeiro – UFRJ, UERJ e CEFET-RJ. Centrais à proposta estão a coleta de dados e a oferta de oficinas aos alunos com intuito de desenvolver uma reflexão crítica acerca da construção cognitiva na qual estão inseridos a partir de um raciocínio científico. Sobre essa ambição, a autora conclui que é “uma ferramenta de metacognição que leva o leitor a examinar o seu próprio modo de pensar o ensino de língua na escola”.

A entrevista com a pesquisadora Katrien Segearts revela igual caráter ambicioso e inédito. A coordenadora do laboratório de Neurociências da Linguagem da Escola de Psicologia da Universidade de Birmingham (Reino Unido) desenvolve estudos sobre o processamento sintático e semântico associado a contextos comunicativos usando paradigmas experimentais inovadores. A equipe trabalha com a interação com avatares, entre interlocutores no scanner da ressonância magnética, e, na mais nova proposta, por jogos comunicativos. Nessa última proposta, os participantes irão jogar enquanto há coleta de EEG em tempo real. Ademais, os seus estudos com idosos geraram atenção na mídia, pois evidenciam a relação entre manutenção de processamento linguístico e a boa forma física. Na entrevista, ela fala sobre interdisciplinaridade, cooperação entre colegas, e experimentação com novos modos de investigar linguagem, mente e cérebro.

Em um estudo com forte controle metodológico, Gustavo Lopez Estivalet discute as “Variáveis

lexicais e ortográficas no acesso lexical das palavras do português brasileiro”. A partir de um experimento de decisão lexical na modalidade visual, ele objetiva investigar a frequência lexical, vizinhança ortográfica, categoria gramatical, número de letras e sílabas como possíveis influências no acesso lexical. O controle dessas variáveis é essencial na verificação de modelos da organização do léxico mental vigentes na literatura. Deste modo, Estivalet segue na sua trajetória de uma série de artigos com intuito e contribuir para um maior rigor na validação e seleção de estímulos lexicais com o uso de corpora na pesquisa em psicolinguística experimental no português brasileiro.

Por fim, as autoras Késsia Henrique e Cristina Name apresentam resultados de um estudo clássico psicolinguístico no artigo “O processamento de múltiplas dependências não adjacentes no PB”. Nele, as autoras investigam o processamento de estruturas com elementos codependentes, marcados morfossintática e/ou fonologicamente, com material interveniente. O possível custo cognitivo para essas dependências não adjacentes do tipo aninhado (ex. O_{A1} carro $A1$ que os $A2$ mecânicos $A2$ consertaram $B2$ bateu $B1$) já tem sido bastante explorado em estudos de aquisição. As autoras contribuem neste estudo, que conta com dois experimentos do tipo tarefa de labirinto (*maze task*), com dados importantes de processamento em adultos. Tempo de leitura maior na condição aninhada sugere que dependências nessa configuração causam maior demanda também no processamento linguístico adulto.

A grande variabilidade de temáticas, metodologias e perspectivas teóricas refletem a grande diversidade e complementaridade dos estudos linguísticos existentes atualmente no Brasil. Há um fio condutor no conjunto apresentado que é a preocupação em gerir, gerar e analisar dados com rigor metodológico, com as mais variadas possibilidades: documentar línguas, falsear hipóteses e investigar fenômenos empíricos no âmbito linguístico. Nós, editoras desta edição, esperamos que vocês, os leitores, encontrem inspiração científica a partir da riqueza dessa pluralidade.

REFERÊNCIAS

ALTMAN, Maria Cristina Fernandes Salles. Os eternos gramáticos: algumas considerações sobre norma e prescrição na descrição linguística. *In Homenagem: 80 anos de Evanildo Bechara*, 2008.

BECKNER, Clay; BLYTHE, Richard; BYBEE, Joan; CHRISTIANSEN, Morten H.; CROFT, William; ELLIS, Nick C.; HOLLAND, John; KE, Jinyun; LARSEN-FREEMAN, Diane; SCHOENEMANN, Tom. Language is a complex adaptive system. *Language Learning*, Ann Arbor, vol. 59, n.1, p. 1-26, dezembro 2009.

BOD, Rens; HAY, Jennifer; JANNEDY, Stefanie. *Probabilistic linguistics*. MIT Press, 2003.

- BLOOMFIELD, Leonard. *Language*. New York: Holt, Rinehart & Winston, 1933.
- BOOIJ, Geert. Construction morphology. *Language and Linguistics Compass*, United Kingdom, v. 4, n. 7, p. 543-555, 2010.
- BYBEE, Joan. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.
- CRUZ, Regina; SEARA, Isabel; MOUTINHO, Lurdes. Intonation of Yes/No question in Portuguese: analysis of the relationship between intonation and lexical stress in the AMPER-POR corpus. In Nebot, Adrián Cabedo (ed.). *Perspectivas actuales en el análisis fónico del habla. Tradición y avances en la fonética experimental*, Valência: Universidade de Valência, p. 35-44, 2015.
- ELLIS, Nick C.; LARSEN-FREEMAN, Diane. *Language as a complex adaptive system*. Oxford: Wiley Blackwell, 2009.
- FERRARI, Lilian. *Introdução à linguística cognitiva*. São Paulo: Contexto, 2014.
- HOLMBERG, Anders. Linguistic Typology. In ROBERTS, Ian. (org.), *The Oxford Handbook of Universal Grammar*, 354–76. Oxford University Press, 2016.
- KAYNE, Richard. *French syntax*. Cambridge, Massachussets: The MIT Press, 1975.
- KRETZSCHMAR JR., William A. *Linguistics and complex systems*. Cambridge: Cambridge University Press, 2015.
- MORAES, João Antônio. A Entoação Modal Brasileira: Fonética e Fonologia. *Caderno de Estudos Linguísticos: Campinas*, v. 25, p. 101-111, 1993.
- MUFWENE, Salikoko S. The emergence of complexity in language. An evolutionary perspective. In: MASSIP-BONET, Àngels; BASTARDAS-BOADA, Albert (eds.), *Complexity perspectives on language, communication and society*. Dordrecht: Springer, 2013. p.197-218.
- MUFWENE, Salikoko S.; COUPÉ, Christophe; PELLEGRINO, François. *Complexity in Language. Developmental and evolutionary perspectives*. Cambridge: Cambridge University Press, 2017.
- SAPIR, Edward *Language: An Introduction to the Study of Speech*. Roxy Harris, Ben Rampton, 1921
- SAUSSURE, Ferdinand de. *Curso de linguística geral*. São Paulo: Cultrix, 1975[1916].

TANENHAUS, Michael K.; SPIVEY-KNOWLTON Michael J.; EBERHARD, Kathleen; SEDIVY, Julie C. Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634, jun 1995.

TORN-LEESIK, Reeli; VIJA, Maigi. Acquisition of the Impersonal Voice by an Estonian Child. *Journal of Baltic Studies*. v. 43, n. 2, 251-271, 2012. Disponível em <https://doi.org/10.1080/01629778.2012>.

**THE NEW NEUROCOGNITIVE SCIENCE OF LANGUAGE: INNOVATIVE
EXPERIMENTAL DESIGN FOR STUDYING SYNTACTIC AND SEMANTIC
PROCESSING IN COMMUNICATIVE SETTINGS, AN INTERVIEW WITH KATRIEN
SEGAERTS**

Ana Regina Calindro¹

Marije Soto²



Katrien Segaert is the head of the Neuroscience of Language lab at the School of Psychology at the University of Birmingham (UK), where she is also affiliated with the Centre for Human Brain Health. Her research deals with the major question of how syntax and semantic processing are instantiated in the brain within the context of communicative settings, and how the neurobiological circuitry for sentence processing changes throughout our lives. Her lab does experimental studies with a variety of techniques, ranging from EEG and fMRI to classical psycholinguistic methodologies. Her work has been really innovative in the ways she has managed to bring communicative context to experimental paradigms: participants engaged with avatars, or interacted with interlocutors while lying in the MRI scanner. Another series of her studies that have attracted some media attention are those involving aging adults: results have shown surprising correlations between language processing and fitness. She has agreed to talk to us about working with novel experimental methodologies in a

1 Profa. Adjunta do Departamento de Linguística e Filologia, Faculdade de Letras, Universidade Federal do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: anacalindro@letras.ufrj.br.

2 Profa. Adjunta do Departamento de Estudos da Linguagem, Instituto de Letras, Universidade do Estado do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: marijesoto@hotmail.com.

highly interdisciplinary field.

A/M: Because you work in an interdisciplinary field, perhaps your research is not that well known to people who work with more traditional linguistics. So, could you describe for those that may not be so familiar with cognitive neuroscience and psycholinguistics, what it is that you do?

K: I look at language not from a linguistic perspective, it's a psycholinguistic perspective. So it's all about behavior. It's how people behave, how they produce and comprehend language. For example - if we're talking about syntax - which syntactic structure do they produce, how fast do they produce it, are they influenced by the structure they just heard? It's always an empirical question. Then I try to relate that to theory, in a sense that you wonder, what does the data you have tell you about theories? For example, how is this syntactic information represented in our cognitive space or in the brain?

A/M: What is your view on the nature of language? This goes perhaps a little bit more back to theories, and how it affects the way you choose subject matters, methodologies?

K: It sounds strange talking to two linguists, but I don't think I am ever led, in terms of my research question, by a theory. I suppose I am always driven by something I want to know, and it's about knowing if it is this correct or not? The starting point is always an empirical question, indirectly, it is driven by theory, because, of course, there are previous people who have thought about these questions. But the theory itself is not the starting point. The starting point is something we don't know yet, because there's more than one theory, or the theories aren't specified enough. I also don't feel tied to a specific method ever. I just have a question and I try to figure out what the best method to answer it would be. It could be a very simple behavioral experiment or a super complicated fMRI experiment. It doesn't matter as long as it answers the question.

A/M: Can you give an example? Maybe something to illustrate it?

K: One line of research that I did quite a lot of experiments on, was about syntactic processing: how we put the building blocks together when we're comprehending language, versus when we're producing language. There are a lot of theories about how we do this in comprehension and there are a lot of theories about how we do this in production. I started from scratch. I thought we need to think about how we do this when we're listening to language versus when we're speaking. Maybe there's more shared than there are differences, so maybe the starting point of all of it is wrong. Then I did a bunch of experiments to see if there is something shared in the brain when you're engaging

in syntactic processing. It turned out that you had all the exact same brain regions, the exact same network of brain regions, working together to do syntactic processing irrespective of whether you were listening or speaking.³ So if I had started out from a specific theory and tested a theory, I would have never gotten there.

A/M: On the other hand, I know people who are not at all interested in looking at syntax, for example. So, your view of the nature of language obviously did lead you to that point, because there are people who never feel like investigating syntax. Maybe you can tell us a little bit about that.

K: Yes, that is true and I definitely don't want to say that you should never start from a theory. Even if you don't start from a theory, you always have to get to a point where you reflect on how your findings are relevant for the theories that we have. You always have to link to theory, but I suppose it could be at the start or at the end, or at both.

A/M: For example, why are you interested in syntax as opposed to pragmatics or semantics?

K: [Syntax], I think, was a logical starting point when I was a graduate student. Now, I am starting to try and see the similarities also between what you're doing on a meaning level versus what you're doing on a syntactic level, and there are so many things that are similar. You have these building blocks you have to put together, then you put them together in a way that you are doing a lot more than just putting them next to each other, you'll have to do some sort of process where you're doing much more than just adding them. In my opinion there's a lot there that could be similar for the two [syntax and semantics], as well in terms of how this works in the brain and how this is implemented. The boundaries are very fluid.

A/M: Where does your interest in syntactic priming come from?

K: I'm not sure. I was a master student and I was doing my masters in psychology in Belgium and there weren't really any psycholinguists in the psychology department. and I just thought 'ooh, this might be interesting'. It was only afterwards when I was a graduate student, and I learned much more about it [priming paradigm] that I really understood how powerful this can be, because you can learn so much about how things are represented in your cognitive space, but also in the brain. It's such a useful tool if you're doing neuroimaging research because it allows so much control. It's a really helpful design if you are a neuroimaging person, but I didn't understand that until much later.

³ Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension — An fMRI study. *Cerebral Cortex*, 22, 1662-1670. doi:10.1093/cercor/bhr249. (the article can be accessed here: https://pure.mpg.de/rest/items/item_979636_14/component/file_1481364/content).

A/M: Is your background in psychology?

K: I'm not a linguist at all. I feel very nervous talking to linguists because of it, because there are so many things I don't understand.

A/M: And vice versa!

A/M: A lot of your papers are innovative in that you try to bring communicative settings into experimental paradigms. How do you feel about the challenges of this endeavor, and what constitutes a valid communicative setting minimally?

K: I guess a minimal communicative setting is when there is a message that's exchanged between two interlocutors, successfully or not successfully. You have a speaker, who has a message in mind, then needs to perform all these processes to get some sort of motor output and speech. Then you have a listener, all of that [which was produced by the speaker] reaches their auditory cortex, they have to do all of these processes, and verify if this is the message that the speaker intended. Well this is a very narrow view on what communication is. Of course, there's sign language, so you could have two signers. Maybe you don't even need two people, it could be a writer and a reader...

A/M: The reason I'm asking this is because there are some linguists who claim we can only investigate language in a highly contextualized setting. What is your view on that?

K: Some of us need to. I don't know if every one of us need to, because there are a lot of questions you can answer when you don't [investigate in a communicative contextualized setting]. But at least some of us need to because we know that languages process differently in a communicative setting versus when there isn't one. I mean you can answer all sorts of super nitpicky questions about how language works, but at some point you have to connect it back to everything around it, because they influence each other. You know that you are going to communicate differently with different people, and that those things matter. It matters if people have aligned with each other previously, whether they've shared information previously, whether there is a social context, what the social context is. Maybe you have a teacher and a student, or two friends, or a husband and a wife. If the husband and the wife just had a fight or not, or if they are having a super good day with each other. You can have someone you love, someone that you don't get along with at all, someone who has a lot of empathy, someone that doesn't. All of those things are going to influence how you produce and how you comprehend language. It's not an isolated process that is completely modular and separate from all of those other things, and, of course, there are a lot of things we need to learn within this module of

language, but at some point some language researchers need to start thinking about how all of these things around it influence language processing.

A/M: We saw in your research that communicative setting is important. How do you emulate real communicative situations in an experimental environment?

K: It's horrible, really difficult, because you can try in so many different ways to make it more naturalistic, but it will never be. Unless you start studying something like this [referring to the actual conversation between interviewers and interviewee], but then you lose experimental control. So you always have to balance the two, and I suppose it depends on your research question how you end up balancing the two.

A/M: Do you think it is valid to compare an experiment, when you have a controlled environment, to a certain real situation that has been recorded, for example?

K: It depends on the research question. Because when you're emulating something that's real, there are so many things that come into play, and, if you add a thousand confounding variables, what have you learned? You still don't really know what you're studying. So you have to then try and add them one by one, but then that's not naturalistic either because you know you're adding them one by one. So it's a really difficult thing to do and we've tried in different ways to make progress with it, and we've learned that it's just really difficult.

A/M: Is it important that it's comparable? Because, as you said, sometimes a very abstract experiment can have a great explanatory power. How do you feel about that? In regards to your research questions, how successful has it been?

K: It takes a lot of experience. It's not something that you get right the first time. We've tried it in many different ways. Simply inviting two participants in the lab and doing a task together. Inviting a participant and a confederate into the lab, where the confederate is a research assistant. We've done it in a virtual reality, where you are interacting with a virtual agent. We've done fMRI experiments⁴

4 fMRI stands for functional Magnetic Resonance Imaging (MRI), which is a technique that measures brain activation indirectly by using the magnetic properties of oxygen bound by hemoglobin is transported in the blood flow to replenish neurons with oxygen after neuronal activity has taken place. Thus, different levels of activation can be inferred in relation to cognitive tasks and/or stimulus presentation in a given experiment by comparing correspondent hemodynamic responses using BOLD measures (blood oxygen level dependent). A classic experimental paradigm in language research, for example, is to record activity in response to sentence stimuli compared to word sequence stimuli; the activity that results from this 'cognitive subtraction' is thought to reflect those operations that are specific to syntactic and semantic sentence building only (as opposed to simply accessing lexical information), and the neuroanatomical areas in which this activation is found are thought to be areas specifically recruited for those operations. (SOTO, 2014)

where someone in the scanner was talking to someone outside of the scanner. We've tried many different things and some have been really successful and some haven't, because sometimes we could replicate the effect that we got the first time. And sometimes when we tried to replicate it the second time it wasn't there anymore, because there are so many different things that influence it and from some of the experiments we've done, we've genuinely learned something, and from others not so much.

A/M: Which kind of paradigms are most successful?

K: In my opinion, we learned a lot from the virtual reality experiments because there you do have a good balance between experimental control and you'll have a communicative context where people interact - a person and a virtual agent. If you have a research question that could be answered with that method, it can work quite well. In one paper, for example, we had people interacting with a virtual agent.⁵ We manipulated this virtual agent depending on facial features and how this agent behaved and spoke. We manipulated how the participant was going to evaluate this virtual agent in terms of: I feel really positive or I feel really negative or I feel sort of average towards this...

A/M: Is that's the social perception you were mentioning in your articles?

K: Yes, that's the one in which the first author is Evelienn Heyselaar. We combined EEG⁶ and virtual reality, so we found that based on the EEG responses you're able to give most attention to someone that you evaluate average. You're able to give less attention to someone that you evaluate really positively or really negatively. That was a really strong and clear result, which also aligns very well with a lot of behavioral experiments we did. We feel confident that attention is probably a mechanism that drives a lot of these social effects, so there we feel like we hit the right balance between a controlled experiment and still having some sort of communicative interaction.

A/M: What was the strangeness factor in that paradigm? How did you manipulate that?

5 Heyselaar, E., Hagoort, P., & Segaert, K. (2016). How social opinion influences syntactic processing—An investigation using virtual reality. *PLoS ONE*, 12(4), Article e0174405.

6 In EEG, brain wave patterns (i.e. their frequencies over time) are measured. Some methodologies analyze frequency modulations in continuous EEG data, other methodologies measure responses to stimuli. In language research, the latter methodology, Event Related Potential (ERP), has a long history since the early 1980s. ERP responses are averaged brain wave responses synchronized to the presentation of a stimuli, such as a written word, or phonological sound that is presented. Averaged responses usually follow a pattern that transcends different languages, modalities, but instead are thought to be a sort of brain signature of a given cognitive process. The N400, for example, is a much replicated response to words presented in isolation or in sentences, that occurs 400ms after stimulus presentation.⁵⁴ (KUTAS; FEDERMEIER, 2011)

K: It turns out that you can manipulate eyebrow movements and blinking rates and all sorts of things where you wonder is this going to have a huge effect? And then you get this super consistent effect on how you feel towards this avatar that you're interacting with, just based on eyebrow movements, blinking rate and speech.

Now we're trying to reflect on the topic of cooperation as a joint action. We're having loads of fun at the moment programming fun communicative games where you have two people who have a challenge, and they need to successfully communicate to complete it. We were inspired by this really funny computer game called "keep talking so nobody explodes". One person has to defuse a bomb and the other person has the bomb manual; the person defusing the bomb can't see manual, so they have to talk about these abstract things, like symbols, they see wires that need cutting, and they have to successfully cooperate and communicate to make sure the bomb doesn't explode. We're going to have very free communication, we have no idea whether this is going work or not. Both participants are going wear EEG caps and we're going to look at oscillatory brain responses.⁷ Then, we will measure the connectivity between the two participants and see if these brain responses align more, that is, whether they synchronize if people communicate successfully.

A/M: There are lots of studies on that with music, aren't there? There was a study here in São Paulo about violinists playing together.⁸

K: There are some studies with a technique called fNIRS⁹, but it doesn't really tell you very well how the brain is working - not as well as with EEG. There are a lot of EEG studies with two participants that are doing nonverbal joint action tasks. We're going to try and bring it to real verbal communication and see. I think we're going to have a lot of fun the next couple of years, but I

⁷ Oscillatory brain responses are the wave-like patterns recorded from neuronal activity. Alternation between negative and positive voltage reflects (post) synaptic activation, which is caused by differently charged environments between neurons through the exchange of neurochemical substances (i.e. neurotransmitters). In connectivity measures, the aim is to see whether brain waves synchronize (i.e. display similar frequencies), for example, in two different brain areas, or in two participants. The supposition is that synchronization of oscillatory brain responses indicate some sort of functional connectivity.

⁸ Vanzella, P., Balardin, J. B., Furucho, R. A., Zimeo Morais, G. A., Braun Janzen, T., Sammler, D., & Sato, J. R. (2019). fNIRS Responses in Professional Violinists While Playing Duets: Evidence for Distinct Leader and Follower Roles at the Brain Level. *Frontiers in psychology*, 10, 164. <https://doi.org/10.3389/fpsyg.2019.00164>

⁹ fNIRS stands for functional Near-Infrared Spectroscopy (fNIR). It is similar to fMRI, in that it takes advantage of blood oxygenation levels in the brain, which vary as neurons increase oxygen demand as a result of engaging in activity thought to reflect cognitive processing. However, instead of the magnetic properties of hemoglobin, fNIR is able to generate functional neuroimages based on the absorption of near infrared light by the hemoglobin that binds oxygen. fNIRS is not as precise as fMRI in spatial resolution, but it has the advantage of being recorded on the scalp, similar to EEG, which leaves more freedom for participants to move or sit up. While temporal precision is better than fMRI, it is not as precise as EEG (SCARAPICCHIA et al., 2017).

don't know if we'll be successful. I don't know whether we'll get it right, the balance between the experimental control and the naturalistic communication. We won't know until three, four years from now. It might be an enormous failure, or a huge success.

So, in these experiments, we are very much interested in looking at cooperation. Our first step is also to look at how empathy affects it. If you're talking about cooperation, this is a really obvious thing. We'll add a lot of empathy questionnaires and we'll see how that influences things.

A/M: What were some of the effects of the artificiality that you have in emulating these communicative settings?

K: It's a difficult question. I imagine it also depends on which experiment and whatever is relevant for the research questions. Maybe you first need to know how it works in a very artificial context with just one person and then you can build from there. I mean, when I say all these other things have an influence on it, I mean they're going to modulate how it works or slightly impact on how it works, but I don't think you all of a sudden get a completely different mechanism. Syntactic priming is one of those things that people study all the time. There are a million syntactic priming experiments out there, and the vast majority of them are with one participant in a very controlled setting. What we know from all the syntactic priming experiments where you do have two participants and you do have a social context, is that you can slightly modulate the effect, but it's not like all of a sudden the effects appear or disappear.

A/M: We also noticed that you recently published a series of studies on aging and language processing?¹⁰ Why did you become interested in this topic, and what was it like for you to work with a different population than you were perhaps used to?

K: In the beginning, when I was a PhD student and a postdoc, I also only did research on young adults, 20 year-olds who come in, or psychology students who need credits for the course, so they have to come and do experiments. It wasn't until I got a little bit older myself and I started thinking about it. I suppose it goes back to the fact that a lot of my research is really driven by my curiosity of knowing how things work. I guess also, as I was getting a bit older, I started to reflect more about being less self-centered and about the society around me. We have this huge increase in people who

10 To cite just a few of the studies, Katrien talks about: Hardy, S., Segaert, K., Wheeldon, L. (PsyArXiv Preprint). Age-related disruption in the use of lexical information during sentence production, despite preserved syntactic planning. (can be accessed here: <https://psyarxiv.com/xrbta>). Segaert, K., Lucas, S.J.E., Burley, C.V., Segaert, P., Milner, A.E., Ryan, M., Wheeldon, L. (2018) Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific Reports*, 8:6715. doi: 10.1038/s41598-018-24972-

are older, but what do we know about how they process language? I considered that there's a lot we know, but that there's also a lot we don't know. And so it was only six years ago, that I started doing aging research. It is very different to interact with older adults. They are less experienced at doing experiments, but they're also a lot more motivated. They really want to do well, they are very chatty, and they want to stay as long as possible, because they want to help you. They always want to know why you're doing the research. But older adults can get upset if they are not good at a task. They start to worry that maybe they're not good at it and other people are, and whether that means that maybe they're not doing well anymore, in terms of their cognitive health. So I learned I have to make sure that I explain things, that I frame things well, that I make sure I don't emphasize accuracy, I don't come up with tasks that they can't do easily. I program my tasks such that the tasks don't end automatically if the older adults make too many mistakes, because you don't want to upset people. You don't want to make a participant feel that their cognitive health is declining just because they can't do a really stupid experiment in a laboratory. So that's something that I've learned to consider that I never considered before.

A/M: What is your opinion about the importance of diversity in participants, for example, in terms of socio-economic backgrounds?

K: I think it's hugely important, I mean, we can come up with the most amazing theory of how we speak or produce language or retrieve words or whatever, if it doesn't apply to the population, what's the point? If it only applies to first-year psychology students, what have we really learned? And [this is especially important now that] we know that older adults process language differently. So I believe theories need to be able to account for that, and especially nowadays aging research is becoming more relevant than it ever was. Just because older adults are a bigger part of the population now than they used to be, especially in Europe¹¹.

A/M: What have you found? Can you tell us something already?

K: Loads of things. I suppose a lot of people still think that in terms of comprehending language for older adults things don't change much, but in terms of producing language you do see a decline with aging. My research doesn't confirm that at all. There is also decline in terms of comprehending language. With regards to behavior, there's also huge individual variability: one 65-year-old is going

¹¹ This is not so much the case in Brazil, where the median age is 33.3 (CIA world factbook, 2018 estimate), although there is a tendency towards a larger percentage of the population to be over 30, 57,1%-in 2018, compared to 52,4% in 2012 (ibge.gov.br). In the UK, for example, 62.5% was over 30 in 2011, and percentages are even higher in other European countries, such as Germany (ons.gov.uk).

to be completely different from another 65-year-old. One could perform at a level of a 30-year-old while the other could perform maybe at the level of what you would imagine a 90-year-old would be. In terms of aging research, there's a lot of variability that has nothing to do with age. It's a lot more about how you age. That includes things such as genetics, of course, but it also includes a host of lifestyle things, for example, how much you exercise, how much you stay mentally active, your diet, how well you sleep. The list is endless.

A/M: So also could you maybe conclude that these variables modulate the effects to a higher extent in older people than in younger people?

K: Yes! Because they have such a huge impact on how the aging process makes cognition decline. There's such individual variability in how people age. The individual variability in just about any task in a group of 60-year-olds is much more than the individual variability in a group of 20-year-olds. As you age, there's a huge decline in every aspect of brain structure, so your brain needs to reorganize its functional processing, but how well your brain does that is very much dependent on all of these other factors. And that influences cognition, of course, which includes language for me.

A/M: Since this is socially so relevant now, is that something you can get funding for relatively easily?

K: I can't get funding easily for anything! But the funding that I have gotten, let me think carefully, it was for aging research. So maybe that is a reflection of the fact that it is becoming more important. It was not something that was a huge thing to consider about 15 years ago, but now we really have to figure this out, because our society is changing.

A/M: And in terms of methodology, did you have questionnaires that you would use for younger people and then you just applied them to older people?

K: No. I feel I had to learn gradually, you can never just simply have an experiment you did on young people, do the same with older people and see what happens. It's never that simple because they need more time for things, and you want to measure language, not how fast they can [perform a task]. There are some things you have to take into account. For example, when you do a tip-of-the-tongue experiment where you measure how easily people can retrieve words, you can do that by giving people pictures or definitions of things to make them find words. If you devise an experiment for young people, you can ask them who is married to Brad Pitt. But a 70 year old doesn't know about that. They know about actors that were famous 50 years ago. Even the nature of your stimuli, they

need to be different, because they need to be relevant.

A/M: We read that you have a paper that shows that older people have to be fit in order to have better cognitive processing. How do you measure that a person is fit? With a questionnaire?

K: Not a questionnaire. I am not scared of interdisciplinary things. I really enjoy working with people who know things that I don't. And for this [study] I worked with people who are exercise physiologists; they measure how fit people are. So basically they put people on a stationary exercise bike, which gets incrementally more difficult. The resistance gets bigger on the exercise bike, and they have to exercise, while we measured oxygen uptake. So they were wearing some sort of masks, and that measured oxygen uptake, while they were exercising. And they had to keep exercising until they reached 80% of their maximal heart rates because it would be unsafe to go further than that. From that you can derive a super objective measure of how fit someone is. If you go to a fitness center right now and they want to devise an exercise plan for you, that's what they would do. I believe it's really important to view being a researcher as something that is really interactive and to not be scared by things you don't know because that's the whole point of being a researcher, isn't it?

A/M: The next question is related to this topic of interdisciplinarity. In terms of methodology, there's such great technology out there. But at the same time it's limiting, perhaps? Because the experimental environment restricts how natural participants can act, wouldn't you agree?

K: One of the biggest challenges is that language comprehension is a lot easier to study than language production. That's an issue of experimental control first. Because if you study language comprehension, you make people listen or read. You have full experimental control. If you study language production, you have to find intelligent tasks to make people produce language, but then in a way that you want them to, so that you can do something with the data. So it's an issue of devising good tasks, but it's also an issue of the actual neuroimaging methodology because basically for every single neural imaging methodology, when people are speaking, your signal gets incredibly affected and really really messy. So with fMRI, which is all about measuring blood flow in the brain, you're lying in an MRI scanner, if you're speaking, you're moving not just your jaw but also your head. But if your head is moving one millimeter, this ruins your signal, and because you're speaking, you have, what they call susceptibility artifacts, which is basically air. Air is coming out of your mouth when you're speaking. Speaking also messes up the signal for EEG, and for MEG as well. Speaking means movement. You're measuring electrical activity here [points to temples], and the electrical activity that you pick up is a lot weaker than the noise that is coming from your jaw and your muscles. The

muscle artifacts are huge and your electrical signal because of language processing is very small. So it's an issue of tasks and an issue of the neuro imaging methodology. Production's really hard.

Again, it's about finding people who can help you. When we did production research in the MRI scanner, we worked together with people who were MRI physicists. Now I'm trying to do production with EEG and I work together with someone who's really strong in MEG methodology and has been for many years. I could never figure this out on my own.

A/M: How do you view your position as a researcher coming from humanities in a highly interdisciplinary field in which statistics, technology, computation and biology are combined into one? Is it difficult to understand all these fields?

K: I try to just constantly remind myself that it's not my responsibility to know all of these things. As much as I can, I feel I have to be strong in terms of research design, and statistics, also I have to know the literature, so that I can ask relevant research questions, but it's not part of my job to know the details of every single neuro imaging methodology. I need to know some of it and then find someone to help me. It's true statistics are important and I think that that is an advantage I have as a psychologist. If you do an undergraduate and a master's in psychology, it means years and years of training in statistics. So sure, statistics are challenging in a sense that you also need to keep updating your knowledge because things change. I'm not sure I see it as a limiting factor or something that makes things harder. I guess I see it more as an opportunity than a constraint, because there's so much that's possible. But then, I'm privileged of course because I have always worked in a place with very intelligent colleagues. The most important thing is to surround yourself with colleagues and to see it as something that's collaborative. You need to have good working relationships with people. Help each other out.

A/M: It's also a matter of infrastructure perhaps? I don't know what you've noticed about Brazil, [Katrien has visited Brazil several times] for example.

K: Yes, I know. That's why I said I feel that I'm very lucky. When I went to Florianópolis and Mailce [Mailce Borges Mota, who is a full professor at UFSC] wanted to do EEG. So we helped her set up the lab and showed her how to collect data. If you don't have people who can help you, you can't get anywhere. But you don't necessarily need to have people right there in the same place with you. You can have people anywhere in the world who can help you.

A/M: We're coming to the final few questions. This is a more general, but very relevant,

question: how far has cognitive neuroscience come in terms of understanding the functional neural architecture in relation to language processing?

K: I don't know. For a technology that hasn't been around for that long, if you think about it, a lot of amazing stuff has happened. So much amazing stuff has happened in that small period of time, that in the next 50 years... I mean, by the time I am retired, I'm sure I will understand *nothing* of what the people are doing. But that is an optimistic thought, right? 40, 50 years from now we'll have learned so much, that I won't understand things. Now, we know about 1% of what goes on in the brain!

But it's also really exciting there are so many things we can do. You'll never run out of projects or research questions or experiments, I'm sure I can keep doing this until I retire.

I feel that language is sometimes the ugly duckling of cognition. You have your perception, attention and working memory, and it's all so cool - and the latest methods and this and that. Language is often like the ugly duckling, running a bit behind. But I think language is doing really well at the moment, and I think we should be really happy to be in this field

A/M: We saw your research has had some attention in the media. What is your view on popularizing neuro cognitive science especially with regards to language? What are the challenges of translating complicated scientific concepts procedures and analyses etc.

K: I don't think *all* research needs to be popularized, not all research needs to end up in the newspaper. But I am of the opinion that if you *do* do something that is relevant for people to know about, for example, if there are many reasons that exercising is good for you [i.e. it has a positive effect on language functioning], you do have a duty to spend some time communicating it. Because otherwise you're just a researcher in your office with a closed door. In the same way, it's your duty to talk to your colleagues, talk to people. Some studies you can really easily explain to your neighbor and other studies not so much. So I figure that if it's something that you want to talk to your neighbor about, it's really easy to talk to a journalist about it as well.

REFERENCES

KUTAS, Martha; FEDERMEIER, Kara. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol.* 62:621-47, 2011. doi: 10.1146/annurev.psych.093008.131123.

HARDY, Sophie M.; SEGAERT, Katrien; WHEELDON, Linda. (PsyArXiv Preprint). Age-related disruption in the use of lexical information during sentence production, despite preserved syntactic planning.

SCARAPICCHIA, Vanessa; BROWN, Cassandra; MAYO, Chantel; GAWRYLUK, Jodie R. Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies. *Frontiers in human neuroscience*, 11, 419, 2017. <https://doi.org/10.3389/fnhum.2017.00419>

SEGAERT, Katrien; MENENTI, Laura; WEBER, Kirsten; PETERSSON, Karl Magnus; HAGOORT, Peter. Shared syntax in language production and language comprehension — An fMRI study. *Cerebral Cortex*, 22, p. 1662-1670, 2012, doi:10.1093/cercor/bhr249.

SEGAERT, Katrien; LUCAS, Sam .J.E.; BURLEY, Claire V.; SEGAERT, Pieter, MILNER, A.E.; RYAN, Matthew, WHEELDON, Laura. Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific Reports*, 8:6715. doi: 10.1038/s41598-018-24972, 2018.

SOTO, Marije. *ERP and fMRI evidence of compositional differences between linguistic computations for words and sentences* Orientadora: Anieli Improta França. Coorientador: Aline Gesualdi Manhães. 2014. 185 f. Tese (Doutorado) – Programa de Pós-graduação em Linguística, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2014.

VANZELLA, Patrícia; BALARDIN, Joana B.; FURUCHO, Rogério A.; ZIMEO MORAIS, Guilherme Augusto; BRAUN JANZEN, Thenille; SAMMLER, Daniela; SATO, João R. fNIRS Responses in Professional Violinists While Playing Duets: Evidence for Distinct Leader and Follower Roles at the Brain Level. *Frontiers in psychology*, 10, 164. 2019. <https://doi.org/10.3389/fpsyg.2019.00164>

**A NOVA CIÊNCIA NEUROCOGNITIVA DA LINGUAGEM: PARADIGMAS
EXPERIMENTAIS INOVADORES PARA ESTUDAR O PROCESSAMENTO SINTÁTICO
E SEMÂNTICO EM CONTEXTOS COMUNICATIVOS, UMA ENTREVISTA COM
KATRIEN SEGAERTS**

Ana Regina Calindro¹

Marije Soto²



Tradução: Fabrício Gomes Paiva³, Ana Regina Calindro e Marije Soto

Katrien Segaert é a coordenadora do laboratório de Neurociências da Linguagem da Escola de Psicologia da Universidade de Birmingham (Reino Unido) e afiliada ao Centro de Saúde do Cérebro Humano. Sua pesquisa investiga como a sintaxe e o processamento semântico são instanciados no cérebro dentro de contextos comunicativos, e como o circuito neurobiológico para o processamento de sentenças muda ao longo da vida. O seu laboratório faz estudos experimentais com uma pluralidade de técnicas, variando de eletroencefalografia (EEG) e imagem de ressonância magnética funcional (fMRI)⁴ a metodologias psicolinguísticas clássicas. O trabalho da pesquisadora tem se mostrado

1 Profa. Adjunta do Departamento de Linguística e Filologia, Faculdade de Letras, Universidade Federal do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: anacalindro@letras.ufrj.br.

2 Profa. Adjunta do Departamento de Linguística e Filologia, Faculdade de Letras, Universidade Federal do Rio de Janeiro; membro do Programa da Pós-graduação em Linguística, Universidade Federal do Rio de Janeiro; email: marijesoto@hotmail.com.

3 Graduando do curso de Inglês e Literaturas de Língua Inglesa, Universidade do Estado do Rio de Janeiro; email: fabriciogp150@gmail.com

4 fMRI significa Imagem por Ressonância Magnética Funcional, que é uma técnica que mede a ativação cerebral indiretamente, usando as propriedades magnéticas do oxigênio ligadas à hemoglobina e transportada no fluxo sanguíneo para reabastecer os neurônios com oxigênio após a atividade neuronal ocorrer. Assim, diferentes níveis de ativação podem ser inferidos em relação a tarefas cognitivas e/ou apresentação de estímulos em um determinado experimento, comparando respostas hemodinâmicas correspondentes usando medidas BOLD (em português, dependentes do nível de

realmente inovador nas maneiras pelas quais conseguiu associar o contexto comunicativo aos paradigmas experimentais: participantes envolvidos com avatares ou interagindo com interlocutores enquanto estavam no *scanner* de MRI. Seus estudos que envolvem adultos em envelhecimento têm atraído atenção da mídia, pois os resultados mostraram correlações surpreendentes entre o processamento da linguagem e a boa forma física. Katrien concordou em conversar conosco seu trabalho com novas metodologias experimentais em um campo altamente interdisciplinar.

A/M: Como você trabalha em um campo interdisciplinar, talvez sua pesquisa não seja tão conhecida por pessoas que lidam com a linguística mais tradicional. Você poderia descrever o que você faz para aqueles que podem não estar tão familiarizados com a neurociência cognitiva e a psicolinguística,?

K: Eu não olho a linguagem de uma perspectiva linguística, mas sim de uma perspectiva psicolinguística. Deste modo, tudo está relacionado ao comportamento. Como as pessoas se comportam, como produzem e compreendem a linguagem. Por exemplo - se estamos falando sobre sintaxe - que estrutura sintática eles produzem, quão rápido eles a produzem, eles são influenciados pela estrutura que acabaram de ouvir? Sempre parto de uma pergunta empírica. Assim, tento relacionar isso à teoria, no sentido de que eu penso sobre o que os dados que tenho em mãos podem dizer a respeito das teorias? Por exemplo, como essas informações sintáticas estão representadas no nosso espaço cognitivo ou no cérebro?

A/M: Qual é a sua opinião sobre a natureza da linguagem? Talvez isso se relacione um pouco mais com as teorias, ou seja, como certos arcabouços teóricos afetam a maneira como você escolhe os tópicos de pesquisa, as metodologias?

K: É estranho conversar sobre isso com duas linguistas, mas acho que nunca sou levada à minha pergunta de pesquisa por uma teoria. Creio que sou sempre impulsionada por algo que eu queira saber, se algo é verdadeiro ou não. O ponto de partida é sempre uma questão empírica. Indiretamente, ela é impulsionada pela teoria, porque, claro, houve outras pessoas que pensaram a respeito dessas questões. Porém, a teoria em si não é o ponto de partida. O ponto de partida é algo que não sabemos ainda, porque há mais de uma teoria, ou as teorias não são suficientemente especificadas. Também não me sinto ligada a um método específico. Eu só tenho uma determinada pergunta e tento imaginar

oxigênio no sangue). Um paradigma experimental clássico na pesquisa de linguagem, por exemplo, é registrar a atividade em resposta a estímulos de sentenças em comparação aos estímulos de sequência de palavras; acredita-se que a atividade resultante dessa 'subtração cognitiva' reflita aquelas operações que são específicas apenas para a construção de sentenças sintáticas e semânticas (em oposição a simplesmente acessar informações lexicais) e as áreas neuroanatômicas nas quais essa ativação é encontrada são consideradas áreas especificamente recrutadas para essas operações (SOTO, 2014)

qual seria o melhor método para respondê-la. Poderia ser um experimento comportamental muito simples ou um experimento de fMRI super complicado. Não importa, contanto que isso responda à pergunta.

A/M: Você pode dar um exemplo? Algo para ilustrar o que disse?

K: Uma linha de pesquisa na qual fiz vários experimentos, foi a de processamento sintático: como colocamos os blocos de construção linguística juntos no processo de compreensão da linguagem comparado ao que fazemos quando estamos produzindo linguagem. Há inúmeras teorias sobre a compreensão da linguagem e produção. Porém, eu comecei do zero. Pensei que precisaríamos refletir sobre o que fazemos quando estamos ouvindo a linguagem *versus* quando estamos falando. Talvez haja mais fatores compartilhados do que diferenças, talvez o ponto de partida de tudo esteja errado. Então, fiz vários experimentos para ver se existe algo compartilhado no cérebro quando você está engajado no processamento sintático. Verifiquei que as mesmas regiões cerebrais, exatamente a mesma rede de regiões cerebrais, estavam trabalhando em conjunto para realizar o processamento sintático independentemente se você estava ouvindo ou falando⁵. Logo, se eu tivesse começado a pesquisa a partir de uma determinada teoria e testado a mesma, eu nunca teria chegado a essa conclusão.

A/M: Por outro lado, há pessoas que não estão interessadas em olhar para a sintaxe, por exemplo. Deste modo, sua visão da natureza da linguagem obviamente a levou a esse ponto, pois há pessoas que nunca pensariam em investigar a sintaxe. Talvez você possa nos contar um pouco sobre isso.

K: Sim, isso é verdade e eu definitivamente não quero dizer que você nunca deva começar a partir de uma teoria. Mesmo se você não começar a partir de uma teoria, você sempre tem que chegar a um ponto no qual pode refletir de que maneira suas descobertas são relevantes para as teorias que temos. Você sempre tem que ligar seus resultados à teoria, mas suponho que poderia ser no início ou no fim, ou em ambos.

A/M: Por exemplo, por que você estaria interessada em sintaxe, em oposição a pragmática ou semântica?

K: Acredito que a [sintaxe] foi um ponto de partida lógico quando eu era estudante de graduação. Agora, estou começando a experimentar e ver as semelhanças também entre o que está sendo feito em nível de significado *versus* o que está sendo feito em nível sintático e há tantas coisas que são

5 Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension — An fMRI study. *Cerebral Cortex*, 22, 1662-1670. doi:10.1093/cercor/bhr249. (o artigo pode ser acessado aqui: https://pure.mpg.de/rest/items/item_979636_14/component/file_1481364/content).

semelhantes. Há esses blocos de construção linguística e você tem que colocá-los juntos e de uma forma que seja muito mais do que apenas juntá-los lado a lado. Você terá que engajar em algum tipo de processo no qual você está fazendo muito mais do que apenas somá-los. Eu sou da opinião de que há muitas semelhanças entre as duas [sintaxe e semântica], bem como em termos de como isso funciona no cérebro e como isso é implementado. Os limites são muito fluidos.

A/M: De onde surge seu interesse pela *priming* sintático⁶?

K: Eu não tenho certeza. Quando eu estava fazendo meu mestrado em psicologia na Bélgica e não havia nenhum psicolinguista no departamento de psicologia, eu pensei ‘oh, isso pode ser interessante’. Foi só depois, no doutorado, que aprendi muito mais sobre isso [paradigma de *priming*] que realmente entendi o quão poderoso isso pode ser, uma vez que podemos aprender muito a respeito de como as coisas são representadas no nosso espaço cognitivo, e também no cérebro. É uma ferramenta muito útil, se estiver fazendo pesquisa de neuroimagem, pois permite muito controle. É um design bastante útil. Mas só entendi isso depois de um bom tempo.

A/M: A sua formação é em psicologia?

K: Eu não sou linguista de forma alguma. Por isso fico muito nervosa conversando com linguistas, porque há muitas coisas que eu não entendo.

A/M: E vice-versa!

A/M: Muitos de seus artigos são inovadores, na tentativa de trazer contextos comunicativos para paradigmas experimentais. Como você se sente sobre os desafios deste empreendimento? O que você considera que constitui um ambiente comunicativo minimamente válido?

K: Eu acho que um contexto comunicativo mínimo é quando há uma mensagem trocada entre dois interlocutores, com ou sem êxito. Há/Existe um falante, que tem uma mensagem em mente, e então precisa realizar diversos processos para produzir algum tipo de *output* motor e a fala. Então, há um ouvinte. Tudo (o que foi produzido pelo falante) atinge o seu córtex auditivo. Eles têm que engajar em todos esses processos e verificar se esta é a mensagem que o falante pretendia. Bem, esta é uma visão muito estreita sobre o que é comunicação. Obviamente, há a língua de sinais, então você

⁶ Priming é um fenômeno, muito usado em experimentos psicolinguísticos, no qual exposição a um determinado estímulo (ex. uma palavra ou um tipo de estrutura) influencia a resposta a um estímulo subsequente, ou influencia a chance de a pessoa usar uma determinada estrutura. Por exemplo, ao se mostrar a palavra cachorro, o reconhecimento de uma palavra associada semanticamente como ração vai ser mais fácil, ou quando você é exposto a sentenças passivas, a chance de você logo em seguida produzir uma estrutura passiva é maior. O processo de priming ocorre automática e inconscientemente. Além disso, supõe-se que reflete processos e organização cognitiva mental. (SOTO, 2014)

poderia ter duas pessoas que sinalizam. Talvez, você não precise mesmo de duas pessoas, poderia ser um escritor e um leitor...

A/M: A razão pela qual estou perguntando é porque existem alguns linguistas que afirmam que somente podemos investigar a linguagem em um contexto altamente contextualizado. Qual a sua opinião sobre isso?

K: Alguns de nós precisam. Não sei se todos nós, pois há muitas perguntas que você pode responder quando não precisa. Mas pelo menos alguns de nós precisam, porque sabemos que linguagem é processada de maneira diferente em um ambiente comunicativo, e quando não há [um contexto comunicativo]. Quero dizer, você pode responder a todos os tipos de perguntas super detalhadas sobre como a linguagem funciona, mas em algum momento você precisa conectá-la a tudo ao seu redor, porque o contexto e a linguagem se influenciam mutuamente. Você sabe que vai se comunicar de forma diferente com pessoas diferentes e que essas coisas importam. Faz diferença se as pessoas se alinharam anteriormente, se já haviam compartilharam informações, se existe um contexto social e qual é esse contexto social. Talvez seja um professor e um aluno, ou dois amigos, ou marido e esposa. Se o marido e a esposa brigaram ou não, ou se eles estão tendo um dia incrível um com o outro. Pode ser alguém que você ama; alguém com quem você não se dá bem; ou alguém com quem tenha muita empatia, ou não. Todas essas coisas irão influenciar a forma como você produz e compreende linguagem. Não é um processo isolado, completamente modular e separado de todas essas outras coisas, e, é claro, são vários fatores que precisamos aprender dentro deste módulo de linguagem, mas em algum momento os pesquisadores de linguagem precisam começar a pensar em como todas essas coisas ao redor influenciam o processamento da linguagem.

A/M: Vimos em sua pesquisa que o ambiente comunicativo é importante. Como você emula situações comunicativas reais em um ambiente experimental?

K: É horrível, muito difícil, porque você pode tentar de muitas maneiras diferentes tornar a comunicação mais natural, porém ela nunca será. A menos que você comece a estudar algo assim [referindo-se à conversa real entre entrevistadoras e entrevistada], contudo o controle experimental será perdido. Então, você sempre precisa equilibrar os dois, e suponho que esse equilíbrio depende da sua pergunta de pesquisa.

A/M: Você acha válido comparar um experimento, no qual você tem um ambiente controlado, com uma situação real, por exemplo?

K: Depende da pergunta de pesquisa. Porque quando você está emulando algo que é real, há inúmeras coisas que entram em jogo e, se você adicionar milhares de variáveis incontroláveis, o que você aprendeu? Você ainda não sabe o que está estudando. Então, você deve fazer várias tentativas e adicionar as variáveis uma a uma. Porém, isso também não é natural, pois você sabe que as está adicionando uma a uma. Logo, é algo muito difícil de fazer. Já tentamos de várias maneiras avançar nessa questão, e aprendemos que é realmente muito difícil.

A/M: É importante que seja comparável? Pois, como você disse, algumas vezes um experimento muito abstrato pode ter um grande poder explicativo. O que você pensa a respeito? Em relação às suas perguntas de pesquisa, você acha que tiveram êxito na comparação?

K: É preciso muita experiência. Não é algo que você acerta na primeira vez. Tentamos fazê-lo de muitas maneiras diferentes: simplesmente convidando dois participantes para fazer uma tarefa juntos no laboratório; ou convidando um participante e um assistente de pesquisa para realizar um experimento no laboratório. Criamos uma realidade virtual, onde você está interagindo com um agente virtual. Fizemos experimentos de fMRI, nos quais alguém no *scanner* estava conversando com alguém fora do *scanner*. Experimentamos muitas coisas diferentes. Algumas tiveram muito sucesso outras não, porque às vezes até conseguíamos replicar o efeito que tivemos na primeira vez, mas, quando tentamos replicá-lo uma segunda vez, não encontramos mais o efeito, pois há inúmeras coisas diferentes que o influenciam. A partir de alguns dos experimentos que fizemos, realmente aprendemos alguma coisa e, a partir de outros, nem tanto.

A/M: Quais tipos de paradigmas são mais bem sucedidos?

K: Na minha opinião, aprendemos muito com os experimentos de realidade virtual, porque você tem um bom equilíbrio entre controle experimental e um contexto comunicativo em que as pessoas interagem - uma pessoa e um agente virtual. Se você tem uma pergunta de pesquisa que possa ser respondida com aquele método, ele poderá funcionar muito bem. Em um dos estudos, por exemplo, tínhamos pessoas interagindo com um agente virtual⁷. Manipulamos esse agente virtual a partir de características faciais e de como ele se comportava e falava. Nós manipulamos como o participante avaliaria esse agente virtual em termos de: me sinto muito positivo ou negativo ou me sinto um tanto normal em relação a isso ...

A/M: Essa é a percepção social que você menciona em seus artigos?

7 Heyselaar, E., Hagoort, P., & Segaert, K. (2016). How social opinion influences syntactic processing—An investigation using virtual reality. *PLoS ONE*, 12(4), Article e0174405.

K: Sim, naquele em que a primeira autora é Evelienn Heyselaar. Como combinamos EEG⁸ e realidade virtual, descobrimos que, com base nas respostas do EEG, você consegue dar mais atenção a alguém que avalia como mediano [i.e. que você não tem nem sentimentos fortes nem fracos]. Você consegue dar menos atenção a alguém que você realmente avalia positiva ou negativamente. Esse foi um resultado bastante forte e claro, que também se alinha muito bem com muitos experimentos comportamentais que fizemos. Estamos confiantes de que a atenção é provavelmente um mecanismo que gera muitos desses efeitos sociais; portanto, sentimos que alcançamos o equilíbrio certo entre um experimento controlado e, ao mesmo tempo, ter algum tipo de interação comunicativa.

A/M: Qual foi o fator de estranheza nesse paradigma? Como você lidou com isso?

K: Surpreendentemente é possível manipular os movimentos das sobrancelhas e quantidade de piscadas e todo tipo de coisa. Daí você pensa isso vai ter um efeito enorme? E, então, você obtém esse/um efeito super consistente de como alguém se sente em relação ao avatar com o qual está interagindo, apenas com base nos movimentos da sobrancelha, na quantidade de piscadas e na fala.

No momento estamos tentando refletir sobre a ideia de cooperação como uma ação conjunta. Estamos nos divertindo muito programando jogos comunicativos divertidos, nos quais duas pessoas têm um desafio e precisam se comunicar com sucesso para concluí-lo. Nós fomos inspirados por um jogo de computador muito engraçado chamado “continue falando para que ninguém exploda” (*keep talking so nobody explodes*). Uma pessoa tem que desarmar uma bomba e a outra pessoa tem o manual da bomba; a pessoa que desarma a bomba não pode ver o manual, então eles têm que falar sobre coisas abstratas, como símbolos. Eles veem fios que precisam ser cortados, precisam cooperar e se comunicar com sucesso para garantir que a bomba não exploda. Vamos ter uma comunicação muito livre. Não temos ideia se isso vai funcionar ou não. Ambos os participantes vão usar toucas de EEG e examinaremos as respostas cerebrais oscilatórias⁹. Mediremos a conectividade entre os dois

8 Em EEG, os padrões de ondas cerebrais (ou seja, suas frequências ao longo do tempo) são medidos. Algumas metodologias analisam modulações de frequência em dados contínuos de EEG, outras metodologias medem respostas a estímulos. Na pesquisa de linguagem, a última metodologia, Potencial Relacionado ao Evento (Event Related Potential (ERP)), possui uma longa história desde o início dos anos 80. As respostas do ERP são respostas médias das ondas cerebrais sincronizadas com a apresentação de um estímulo, tal como uma palavra escrita ou som fonológico que é apresentado. As respostas médias, em geral, seguem um padrão que ocorre independente da língua ou modalidade testada, pois, são consideradas uma espécie de assinatura cerebral de um determinado processo cognitivo. O N400, por exemplo, é um sinal neurofisiológico muito replicado, em resposta a palavras apresentadas isoladamente ou em sentenças, que ocorrem 400ms após a apresentação do estímulo. (KUTAS; FEDERMEIER, 2011)

9 Respostas cerebrais oscilatórias são os padrões ondulatórios registrados a partir de atividade neuronal. A alternância entre voltagem? negativa e positiva reflete a ativação (pós) sináptica, a qual é causada por ambientes ao redor dos neurônios que apresentam cargas diferentes através da troca de substâncias neuroquímicas (isto é, neurotransmissores). Nas medidas de conectividade, o objetivo é verificar se as ondas cerebrais são sincronizadas (isto é, exibem frequências semelhantes), por exemplo, em duas áreas cerebrais diferentes ou em dois participantes. A suposição é que a sincronização das respostas

participantes e veremos se essas respostas cerebrais se alinham mais, ou seja, sincronizam-se, caso as pessoas se comuniquem com sucesso.

A/M: Existem muitos estudos sobre isso com música, não é mesmo? Foi feita uma pesquisa aqui em São Paulo sobre violinistas tocando juntos¹⁰.

K: Há alguns estudos com uma técnica chamada fNIRS¹¹, mas isso realmente não diz muito sobre como o cérebro está funcionando - não tanto quanto com o EEG. Existem vários estudos de EEG com dois participantes que realizam tarefas de ação conjunta não verbal. Então, vamos tentar introduzir comunicação verbal real e ver o que acontece. Acho que vamos nos divertir muito nos próximos anos, mas não sei se teremos sucesso. Não sei se acertaremos o equilíbrio entre o controle experimental e a comunicação natural. Só vamos saber daqui a três, quatro anos. Pode ser um enorme fracasso, ou um enorme sucesso.

Portanto, nessas experiências, estamos muito interessados em buscar cooperação. Nosso primeiro passo também é analisar como a empatia afeta. Se você está falando sobre cooperação, isso é algo realmente óbvio. Nós adicionaremos muitos questionários de empatia e vamos ver como isso influencia as coisas.

A/M: Qual é um possível efeito de artificialidade, que se pode ter ao reproduzir esses ambientes comunicativos?

K: É uma pergunta difícil. Eu acredito que também depende de qual experimento e do que for relevante para as questões da pesquisa. Talvez primeiro seja necessário saber como funciona em um contexto muito artificial com apenas uma pessoa e, então, você pode construir [o experimento] a partir daí. Ou seja, quando digo que todas essas outros fatores influenciam [a linguagem], quero dizer que elas modularão o modo como funcionam ou terão pouco impacto sobre o seu funcionamento.

cerebrais oscilatórias indica algum tipo de conectividade funcional.

10 Vanzella, P., Balardin, J. B., Furucho, R. A., Zimeo Morais, G. A., Braun Janzen, T., Sammler, D., & Sato, J. R. (2019). fNIRS Responses in Professional Violinists While Playing Duets: Evidence for Distinct Leader and Follower Roles at the Brain Level. *Frontiers in psychology*, 10, 164. <https://doi.org/10.3389/fpsyg.2019.00164>

11 fNIRS significa Espectrografia de Infravermelho Próximo funcional (fNIR). É semelhante a fMRI, na medida em que tira proveito dos níveis de oxigenação sanguínea no cérebro, que variam conforme os neurônios aumentam a demanda de oxigênio como resultado do envolvimento em atividades que refletem o processamento cognitivo. No entanto, em vez das propriedades magnéticas da hemoglobina, fNIR é capaz de gerar neuroimagens funcionais com base na absorção de luz infravermelha próxima pela hemoglobina que liga o oxigênio. fNIRS não é tão preciso quanto fMRI na resolução espacial, mas tem a vantagem de ser gravado no couro cabeludo, semelhante ao EEG, o que deixa mais liberdade para os participantes se moverem ou sentarem. Embora a precisão temporal seja melhor que fMRI, ela não é tão precisa quanto o EEG (SCARAPICCHIA *et al.*, 2017).

Mas não acho que, de repente, você consiga um mecanismo completamente diferente. O *priming* sintático é uma daquelas coisas que as pessoas estudam o tempo todo. Há um milhão de experimentos de *priming* por aí e a grande maioria deles é feita com um participante em um ambiente muito controlado. O que sabemos de todos os experimentos de *priming* sintático em que você tem dois participantes e um contexto social é que você pode modular um pouco o efeito, mas não é que os efeitos aparecem ou desaparecem de repente.

A/M: Observamos também que você publicou recentemente uma série de estudos sobre envelhecimento e processamento de linguagem?¹² Por que você se interessou por esse tópico e como foi trabalhar com uma população diferente da que estava acostumada?

K: No começo, quando eu era doutoranda e pós-doutoranda, eu também fazia pesquisas apenas com jovens adultos, de 20 anos, ou estudantes de psicologia que precisavam de créditos para o curso e precisavam vir para o laboratório para fazer experimentos. Só quando fiquei um pouco mais velha, comecei a pensar sobre isso. Suponho que tenha a ver com o fato de que muitas das minhas pesquisas são realmente dirigidas pela minha curiosidade de saber como as coisas funcionam. Também acho que, à medida que envelheci, comecei a refletir mais sobre ser menos egoísta e acerca da sociedade ao meu redor. Temos um grande aumento de pessoas mais velhas, mas o que sabemos sobre como elas processam a linguagem? Eu considerei que já sabemos muito, mas que ainda há muito que não sabemos. Comecei a pesquisar sobre envelhecimento há apenas seis anos. É muito diferente interagir com adultos mais velhos. Eles são menos habituados em participar de experimentos, mas também são muito mais motivados. Eles realmente querem se sair bem, são muito falantes e querem ficar o maior tempo possível, porque querem te ajudar. Sempre querem saber por que você está fazendo a pesquisa. Porém, os idosos podem ficar chateados, se não forem bons em uma tarefa. Eles começam a se preocupar que talvez não sejam bons e que outras pessoas sejam, e se isso pode significar que talvez não estejam mais tão bem, em termos de seu desempenho cognitivo. Então, aprendi que tenho que me certificar de que eu explico as coisas direito, contextualizo bem. Tenho o cuidado de não enfatizar a acurácia, ou seja, me certifico de não desenvolver tarefas que eles não possam fazer facilmente. Eu não programo para que minhas tarefas terminem automaticamente, caso os idosos cometam vários erros, pois você não quer deixar as pessoas chateadas. Você não quer fazer com que um participante sinta que seu desempenho cognitivo está em declínio apenas porque não consegue fazer um experimento

12 Para citar alguns dos estudos, Katrien fala sobre: Hardy, S., Segaert, K., Wheeldon, L. (PsyArXiv Preprint). Age-related disruption in the use of lexical information during sentence production, despite preserved syntactic planning (pode ser acessado aqui: <https://psyarxiv.com/xrbta>). Segaert, K., Lucas, S.J.E., Burley, C.V., Segaert, P., Milner, A.E., Ryan, M., Wheeldon, L. (2018) Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific Reports*, 8:6715. doi: 10.1038/s41598-018-24972-

bobo em laboratório. Então, isso é algo que aprendi a considerar e eu nunca tinha considerado antes.

A/M: Qual é a sua opinião sobre a importância da diversidade nos participantes, por exemplo, em termos de contextos socioeconômicos?

K: Eu acho que é extremamente importante, pois podemos inventar a teoria mais incrível de como falamos ou produzimos linguagem, acessamos palavras ou qualquer outra coisa, se ela não se aplica à população, qual é o objetivo? Ou se apenas se aplica aos estudantes do primeiro ano de psicologia, o que realmente aprendemos? E [isso é especialmente importante agora que] sabemos que os idosos processam a linguagem de maneira diferente. Então, acredito que as teorias precisam ser capazes de explicar isso e, especialmente hoje em dia, a pesquisa em envelhecimento está se tornando mais relevante do que nunca. Principalmente porque os idosos são uma parte maior da população agora do que costumavam ser, especialmente na Europa¹³.

A/M: O que você encontrou? Você já pode nos dizer algo?

K: Várias coisas. Quando se pensa em idosos, suponho que muitas pessoas ainda pensam que, em termos de compreensão da linguagem, as coisas não mudam muito. mas em termos de produção de linguagem, você vê um declínio com o envelhecimento. Minha pesquisa não confirma isso de forma alguma. Também existe declínio em termos de compreensão da linguagem. No que diz respeito ao comportamento, há também uma enorme variabilidade individual: uma pessoa de 65 anos será completamente diferente de outra de 65 anos. Uma pode ter um desempenho de uma pessoa de 30 anos, enquanto a outra pode ter um desempenho semelhante ao que você imaginaria que um idoso de 90 anos possa ter. Em termos de pesquisa sobre envelhecimento, há muita variabilidade que não tem nada a ver com a idade em si. Os fatores estão mais relacionados a como você envelhece. Isso inclui coisas como genética, é claro, mas também inclui uma série de aspectos de estilo de vida, por exemplo, o quanto você se exercita, o quanto permanece mentalmente ativo, sua dieta e a qualidade do seu sono. A lista não tem fim.

A/M: Então, você também poderia concluir que essas variáveis modulam os efeitos em maior grau nos idosos do que nos jovens?

K: Sim! Porque eles têm um impacto enorme sobre como o processo de envelhecimento faz com

13 Esse não é o caso do Brasil, onde a idade média é de 33,3 anos (factbook mundial da CIA, estimativa de 2018), embora exista uma tendência de uma porcentagem maior da população ter mais de 30 anos, 57,1% em 2018, comparado a 52,4% em 2012 (ibge.gov.br). No Reino Unido, por exemplo, 62,5% estava acima dos 30 em 2011 e as porcentagens são ainda maiores em outros países europeus, como a Alemanha (ons.gov.uk).

que a cognição diminua. Existe uma variabilidade individual na forma como as pessoas envelhecem. A variabilidade individual em praticamente qualquer tarefa em um grupo de pessoas com 60 anos é muito maior do que em um grupo de pessoas com 20 anos. À medida que você envelhece, há um grande declínio em quase todos os aspectos da sua estrutura cerebral, portanto seu cérebro precisa reorganizar o processamento funcional. Contudo, a qualidade com que o seu cérebro irá fazer essa reorganização depende muito de todos esses outros fatores. Certamente, isso influencia a cognição, o que também inclui a linguagem no meu ponto de vista.

A/M: Como os estudos sobre envelhecimento são socialmente tão relevantes agora, é mais fácil conseguir financiamento para esse tipo de projeto?

K: Não consigo obter financiamento com facilidade para nada! Mas o financiamento que recebi, deixe-me pensar com cuidado, foi para pesquisas sobre envelhecimento sim. Então, talvez isso reflita o fato de que está se tornando mais importante. Não era algo significativo para se considerar há cerca de 15 anos, mas agora temos realmente que pensar sobre isso, porque nossa sociedade está mudando.

A/M: Em termos de metodologia, você tinha questionários que usava para pessoas mais jovens e depois os aplicou a idosos?

K: Não. Eu sinto que tive que aprender gradualmente, você nunca pode simplesmente fazer um experimento com jovens e fazer o mesmo com idosos para ver o que acontece. Nunca é tão simples, porque eles precisam de mais tempo e você quer medir a linguagem, não a rapidez com que eles conseguem [executar uma tarefa]. Há algumas coisas que você deve levar em consideração. Por exemplo, quando você faz um experimento “na ponta da língua” em que mede a facilidade com que as pessoas conseguem recuperar palavras, você pode fazer isso apresentando imagens às pessoas, ou dando definições que os faça encontrar as palavras. Se você planeja um experimento para jovens, pode perguntar a eles *quem é casado com Brad Pitt?* Mas uma pessoa de 70 anos não sabe disso. Eles conhecem atores famosos de 50 anos atrás. Mesmo as características dos seus estímulos precisam ser diferentes, porque precisam ser relevantes.

A/M: Vimos um artigo seu sobre a importância de idosos estarem em forma para ter um melhor processamento cognitivo. Como você mede que uma pessoa está em forma? Com um questionário?

K: Com um questionário não. Não tenho medo de coisas interdisciplinares. Eu realmente gosto de trabalhar com pessoas que sabem coisas que eu não sei. Para este estudo, trabalhei com pessoas

que são fisiologistas do exercício¹⁴ para medir a forma física das pessoas. Então, basicamente, eles colocam as pessoas em uma bicicleta ergométrica estacionária e aumentam gradualmente a dificuldade. A resistência aumenta na bicicleta ergométrica e eles se exercitam, enquanto medimos a capacidade de oxigênio. Eles usavam algum tipo de máscara para medir a captação de oxigênio enquanto se exercitavam. Durante o teste, tiveram que continuar se exercitando até atingir 80% da frequência cardíaca máxima, porque não era seguro ir além disso. A partir disso, você pode obter uma medida super objetiva da forma física de alguém. Se você fosse a uma academia agora, e eles quisessem elaborar um plano de exercícios para você, é isso que eles fariam. Acredito que seja muito importante considerar que ser cientista é algo realmente interativo e não se deve ter medo de coisas que não se sabe, porque esse é o objetivo de ser cientista, não é?

A/M: A próxima pergunta está relacionada a esse tópico da interdisciplinaridade. Em termos de metodologia, existem diversas tecnologias maravilhosas por aí, certo? Mas ao mesmo tempo, limitadoras, talvez? Por que o ambiente experimental restringe a naturalidade com que os participantes podem agir, certo?

K: Um dos maiores desafios é que a compreensão da linguagem é muito mais fácil de estudar do que a produção da linguagem. Essa é, primeiramente, uma questão de controle experimental, certo? Porque se você estuda a compreensão da linguagem, você faz as pessoas ouvirem ou lerem. Há um controle experimental total. Se você estuda a produção da linguagem, precisa encontrar tarefas inteligentes para fazer com que as pessoas produzam linguagem, só que da maneira que você deseja, para que possa fazer algo com os dados. Portanto, é uma questão de desenvolver boas tarefas, mas também é uma questão da metodologia de neuroimagem em si, porque basicamente para qualquer metodologia de neuroimagem, quando as pessoas estão falando, seu sinal é incrivelmente afetado e realmente muito ruidoso. Além disso, com fMRI, que mede o fluxo sanguíneo no cérebro, a pessoa está deitada em um *scanner* de ressonância magnética, se estiver falando, está movendo não apenas a mandíbula, mas também a cabeça. Mas se sua cabeça se mover um milímetro, isso estraga o sinal de fMRI capturado, e, como você está falando, ainda tem o que é chamado de artefatos de suscetibilidade, que são basicamente ar em si, porque o ar sai da sua boca quando você está falando. Falar simplesmente já atrapalha o sinal para o EEG e também para o MEG, pois falar significa movimento. Você está medindo a atividade elétrica aqui [aponta para os ossos temporais logo acima das orelhas] e a atividade elétrica que você capta é muito mais fraca que o sinal ruidoso que sai da

14 “Atua especificamente na mensuração e avaliação de parâmetros fisiológicos, de forma a possibilitar o planejamento de atividades físicas específicas para as necessidades e possibilidades dos clientes” educacaofisica.com.br acesso em 18/05/2020

sua mandíbula e dos seus músculos. Os artefatos musculares são enormes e seu sinal elétrico, devido ao processamento da linguagem, é muito pequeno. Então, é uma questão relacionada às tarefas e à metodologia de neuroimagem. A produção é realmente difícil.

Mais uma vez, é sobre encontrar pessoas que possam ajudar você. Quando fizemos uma pesquisa de produção no *scanner* de ressonância magnética, trabalhamos em conjunto com físicos, especialistas em ressonância magnética. Agora estou tentando fazer a produção com o EEG e trabalho em conjunto com alguém que é bastante forte na metodologia MEG e trabalha com isso há muitos anos. Eu nunca conseguiria descobrir isso sozinha.

A/M: Como você vê sua posição enquanto pesquisadora das ciências humanas em um campo altamente interdisciplinar no qual estatística, tecnologia, computação e biologia fazem parte? É difícil entender todos esses campos?

K: Eu tento me lembrar constantemente de que não é minha responsabilidade saber todas essas coisas., Sinto que tenho que ser forte em termos de projeto de pesquisa e estatística, também tenho que conhecer a literatura, a fim de poder fazer perguntas de pesquisa relevantes, mas não faz parte do meu trabalho conhecer os detalhes de cada metodologia de neuroimagem. Preciso saber um pouco e depois encontrar alguém para me ajudar. É verdade que a estatística é importante e acho que essa é uma vantagem que tenho como psicóloga. Se você faz graduação e mestrado em psicologia, isso significa anos e anos de treinamento em estatística. Mas, claro, a estatística também é desafiadora, no sentido de que você precisa atualizar seu conhecimento, porque as coisas mudam. Não sei se vejo isso como um fator limitante ou algo que dificulta as coisas. Eu acredito que vejo isso mais como uma oportunidade do que como uma restrição, porque há tantas possibilidades. Por outro lado, sou privilegiada, é claro, porque sempre trabalhei em um lugar com colegas muito inteligentes. O mais importante é se cercar de colegas e ver isso como algo colaborativo. Você precisa ter boas relações de trabalho com as pessoas. Ajudar um ao outro.

A/M: Talvez também seja uma questão de infraestrutura? Não sei o que você percebeu sobre o Brasil, [Katrien visitou o Brasil várias vezes] por exemplo.

K: Sim, eu sei. Por isso disse que me sinto muito sortuda. Quando fui para Florianópolis, Mailce [Mailce Borges Mota, professora titular da UFSC] queria fazer EEG. Então, nós a ajudamos a montar o laboratório e mostramos a ela como coletar dados. Se você não tem pessoas que possam te ajudar, você não pode chegar a lugar nenhum. Mas não necessariamente precisa ter pessoas lá no mesmo lugar que você. Você pode ter pessoas em qualquer lugar do mundo que possam te ajudar.

A/M: Estamos chegando às últimas perguntas. Esta é uma pergunta mais geral, porém muito relevante: até que ponto a neurociência cognitiva chegou em termos de compreensão da arquitetura neuro-funcional em relação ao processamento de linguagem?

K: Não sei. Para uma tecnologia que não existe há tanto tempo, se você pensar bem, muita coisa incrível aconteceu. Na verdade, tanta coisa incrível aconteceu nesse pequeno período de tempo, que nos próximos 50 anos, quando eu estiver aposentada, tenho certeza de que não vou entender nada do que as pessoas estão fazendo. Mas esse é um pensamento otimista, não é verdade? Daqui a 40, 50 anos, teremos aprendido tanto, que não vou entender as coisas. Agora, sabemos apenas cerca de 1% do que acontece no cérebro!

Mas também é muito emocionante porque há tantas coisas que podemos fazer. Você nunca ficará sem projetos, questões de pesquisa ou experimentos. Tenho certeza que posso continuar fazendo isso até me aposentar.

Sinto que a linguagem, às vezes, é o patinho feio da cognição. Você tem percepção, atenção e memória de trabalho, e é tudo tão legal - e os métodos mais recentes e isso e aquilo. A linguagem é muitas vezes como o patinho feio, ficando um pouco atrás. Mas acho que a linguagem está indo muito bem no momento e acredito que deveríamos estar muito felizes por estar nesse campo

A/M: Vimos que sua pesquisa teve alguma atenção na mídia. Qual é a sua visão sobre a popularização da ciência neurocognitiva, especialmente com relação à linguagem? Quais são os desafios de traduzir conceitos, procedimentos e análises científicas complicadas, etc.?

K: Não acho que *toda* pesquisa precise ser popularizada, não acho que toda a pesquisa precise acabar no jornal. Mas sou da opinião que se você fizer algo que seja relevante para as pessoas saberem, por exemplo, se há resultados que mostram que fazer exercício é bom para você [ou seja, que isso tem um efeito positivo no funcionamento da linguagem], o pesquisador tem o dever de gastar algum tempo falando a respeito. Caso contrário, você é apenas um pesquisador em seu escritório com a porta fechada. Da mesma maneira, é seu dever conversar com seus colegas, conversar com as pessoas. Alguns estudos podem ser explicados com facilidade ao seu vizinho e outros nem tanto. Então, eu imagino que se é algo que você queira conversar com seu vizinho, é muito fácil conversar com um jornalista sobre isso também.

REFERÊNCIAS

KUTAS, Martha; FEDERMEIER, Kara. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol.* 62:621-47, 2011. doi: 10.1146/annurev.psych.093008.131123.

HARDY, Sophie M.; SEGAERT, Katrien; WHEELDON, Linda. (PsyArXiv Preprint). Age-related disruption in the use of lexical information during sentence production, despite preserved syntactic planning.

SCARAPICCHIA, Vanessa; BROWN, Cassandra; MAYO, Chantel; GAWRYLUK, Jodie R. Functional Magnetic Resonance Imaging and Functional Near-Infrared Spectroscopy: Insights from Combined Recording Studies. *Frontiers in human neuroscience*, 11, 419, 2017. <https://doi.org/10.3389/fnhum.2017.00419>

SEGAERT, Katrien; MENENTI, Laura; WEBER, Kirsten; PETERSSON, Karl Magnus; HAGOORT, Peter. Shared syntax in language production and language comprehension — An fMRI study. *Cerebral Cortex*, 22, p. 1662-1670, 2012, doi:10.1093/cercor/bhr249.

SEGAERT, Katrien; LUCAS, Sam .J.E.; BURLEY, Claire V.; SEGAERT, Pieter, MILNER, A.E.; RYAN, Matthew, WHEELDON, Laura. Higher physical fitness levels are associated with less language decline in healthy ageing. *Scientific Reports*, 8:6715. doi: 10.1038/s41598-018-24972, 2018.

SOTO, Marije. *ERP and fMRI evidence of compositional differences between linguistic computations for words and sentences* Orientadora: Aniela Improta França. Coorientador: Aline Gesualdi Manhães. 2014. 185 f. Tese (Doutorado) – Programa de Pós-graduação em Linguística, Faculdade de Letras, Universidade Federal do Rio de Janeiro, 2014.

VANZELLA, Patrícia; BALARDIN, Joana B.; FURUCHO, Rogério A.; ZIMEO MORAIS, Guilherme Augusto; BRAUN JANZEN, Thenille; SAMMLER, Daniela; SATO, João R. fNIRS Responses in Professional Violinists While Playing Duets: Evidence for Distinct Leader and Follower Roles at the Brain Level. *Frontiers in psychology*, 10, 164. 2019. <https://doi.org/10.3389/fpsyg.2019.00164>

SOBRE A CONSTITUIÇÃO DE CORPORA PARA LÍNGUAS COM POUCOS RECURSOS (*LESS-RESOURCED LANGUAGES*)¹ *ON THE CONSTITUTION OF CORPORA OF LESS-RESOURCED LANGUAGES*

Lílian Teixeira de Sousa²

RESUMO

O uso de *corpora* em estudos linguísticos é bastante antigo, já a área da Linguística de *Corpus* é relativamente nova, tendo sua origem vinculada à ampliação do acesso a computadores e, conseqüentemente, ao Processamento de Linguagem Natural (PLN). À medida que a área foi ganhando influência na pesquisa linguística, o conceito de *corpus* foi se tornando mais específico e elementos como amplitude e referência, além de legibilidade por máquina e tamanho finito, passaram a se tornar fundamentais para a composição de amostras na área. Ao mesmo tempo, no entanto, foram surgindo *corpora* menores e bem menos amplos constituídos com objetivos bastante distintos, como, por exemplo, para a realização de documentação de línguas ameaçadas. Partindo disso, o presente artigo tem por objetivo discutir as diferenças entre *corpora* “prototípicos” criados segundo os pressupostos da Linguística de *Corpus*, e os *corpora* de línguas com pouca presença digital (*less-resourced languages*). Mostro que os *corpora* de línguas com poucos recursos tendem a ser mais especializados e, dificilmente, cumprem todos os critérios exigidos de um corpus amplo e representativo de uma língua. Apesar dos limites impostos por questões específicas de cada língua, concluo que a constituição de *corpora* para línguas com poucos recursos, ainda que não cumpram todos os critérios propostos pela Linguística de *Corpus*, devem ser realizados, e os resultados devem ser aproveitados de diversas formas, seja gerando novas tecnologias, servindo de suporte empírico para teorias linguísticas ou promovendo a língua na comunidade.

Palavras-chave: línguas com poucos recursos, PLN, corpus

ABSTRACT

While the use of corpora in linguistic studies is quite old, Corpus Linguistics is a relatively new area of study, emerging with the expansion of access to computers and, consequently, to Natural Language Processing (NLP). As the subject gained influence within linguistic research, the concept of corpus became more specific. Breadth of sampling and standard references, as well as machine readability and finiteness became essential elements to compose samples. At the same time, however, smaller and much narrower corpora emerged, having distinct purposes, such as those documenting endangered languages. From this understanding, this paper aims to discuss differences between “prototypical” corpora built from the assumptions of Corpus Linguistics and those of less-resourced languages, with a small digital footprint. I demonstrate that the corpora of less-resourced languages tend to be more specialized and hardly ever fulfill the criteria required of a broad and representative corpus. In spite of limitations entailed by issues specific to each language, I conclude that the constitution of corpora for less-resourced languages must be undertaken, even if they do not fulfill all desirable criteria of Corpus Linguistics. The results must be exploited in diverse ways, whether through the creation of new technologies, as empirical support for linguistic theories or in promoting the language in the community.

Keywords: less-resourced languages, NLP, corpus

1 Agradeço aos pareceristas desta revista pelos comentários e sugestões que contribuíram para a qualidade do texto.

2 Universidade Federal da Bahia (UFBA), Setor de Linguística. E-mail: lilian.sousa@ufba.br

1 Introdução

Se considerarmos o conceito de *corpus* como “um conjunto de textos escritos ou falados numa língua, disponível para análise” (TRASK, 2004, p 68-69), pode-se dizer que o uso de *corpora* em estudos linguísticos é bastante antigo, o Corpus Helenístico, por exemplo, data da Antiguidade (BERBER SARDINHA, 2004). Já no século XX, muitos pesquisadores se dedicaram à descrição linguística através de dados coletados para esse propósito. Nos dias de hoje, no entanto, com o uso do computador, a linguística empírica ganhou novos contornos, uma vez que mais pesquisadores passaram a ter acesso ao Processamento de Linguagem Natural³ (PLN) e, conseqüentemente, ampliaram-se as possibilidades de criação e manutenção de *corpora* maiores e mais amplos. Com isso, a Linguística de *Corpus* ganhou influência na pesquisa linguística e passou a ter definições próprias. Berber Sardinha (2004, p 3), por exemplo, a define como uma “abordagem que se ocupa da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística”. Já McEnery e Wilson (2001) tratam a Linguística de *Corpus* não como uma área, mas como uma metodologia que tem por finalidade estudar a língua baseada em exemplos da vida real. Mesmo o conceito de *corpus* se tornou mais específico para dar conta das inovações promovidas pela tecnologia. Segundo Sinclair (2005), um *corpus* é uma coleção de textos de uma língua em formato eletrônico, selecionado por critérios externos para representar, tanto quanto possível, uma língua ou variedade como fonte de dados para a pesquisa linguística.

McEnery e Wilson (2001) tornam o uso do termo *corpus* ainda mais específico ao apresentar quatro características fundamentais, a saber, (i) *amostragem e representatividade (sampling and representativeness)*, um *corpus* deve ter uma amostragem suficiente da língua ou variedade que se quer analisar para obter o máximo de representatividade; (ii) *tamanho finito (finite size)*, um *corpus* é finito independentemente de sua extensão (e.g. em número de palavras); (iii) *formato eletrônico (machine-readable form)*, deve estar em formato que possa ser processado pelo computador; e (iv) *referência padrão (standard reference)*, um *corpus* deve ser uma referência padrão⁴ da variedade de língua que representa. Segundo Aluísio e Almeida (2006), essa última característica é uma diferença marcante entre a concepção de *corpus* para a Linguística e para a Linguística de *Corpus*, já que a construção de um *corpus* nessa perspectiva não serve apenas para uma única pesquisa, mas pode ser

3 Processamento de linguagem natural é uma subárea da ciência da computação, inteligência artificial e da linguística que estuda os problemas da geração e compreensão automática de línguas humanas naturais.

4 Padrão aqui refere-se à regularidade expressa na recorrência sistemática de unidades coocorrentes de várias ordens (lexical, gramatical, sintática etc.)

útil para outros pesquisadores por se tratar de uma referência padrão de uma língua ou variedade, permitindo assim pesquisas sobre qualquer área da linguagem, de fonética a discurso. Assim, uma das principais características da Linguística de *Corpus* é tornar possível generalizações de um *corpus* para uma língua como um todo ou pelo menos para uma variedade ou registro particular.

Paralelo ao desenvolvimento da chamada Linguística de *Corpus*, que, como apontado acima, apresenta definições próprias de *corpora*, começaram a surgir programas de financiamento de projetos de documentação de línguas pouco estudadas ou ameaçadas no mundo todo que resultaram na construção de *corpora* menores e menos amplos (SCANNEL, 2007; OSTLER, 2008; COX, 2011). Isso porque muitos linguistas passaram a reconhecer a necessidade de se realizar a documentação de línguas minoritárias ou ameaçadas. Especialmente depois que Kraus (1992) afirmou que o próximo século veria a morte ou condenação de 90% das línguas humanas. Esses *corpora*, no entanto, não satisfazem todas as condições apontadas no parágrafo acima, já que um bom número de dados presentes nesses bancos foi produzido unicamente para esse fim, não se tratando de dados de contextos comunicativos reais, além do que muitas vezes não conseguem ser amplos o suficiente para serem representativos. Uma outra questão importante é de natureza cultural e ética, já que em sociedades de cultura oral pode haver resistência ao registro escrito e/ou audiovisual⁵, especialmente em rituais sagrados, o que restringe os contextos de documentação.

A questão que surge, então, é sobre a funcionalidade de *corpora* eletrônicos de línguas pouco estudadas, já que, na maioria das vezes, esses *corpora* não apresentam todas as características apontadas como necessárias pelos teóricos da área. Assim, o objetivo desse artigo é discutir as diferenças entre *corpora* “prototípicos” e *corpora* eletrônicos de línguas com pouca presença digital, conhecidas na área de Linguística Computacional como *less-resourced languages* (JOKINEN, 2018; SZYMANSKI, 2011), buscando apontar sua importância e potencial.

O artigo está dividido da seguinte forma: na seção 2, são apresentadas e discutidas algumas definições importantes. Na seção 3, descrevo alguns *corpora* de línguas com poucos recursos e sua funcionalidade. Na seção 4, discuto a importância da criação desse tipo de *corpora* e, por fim, apresento, na seção 5, as Considerações Finais.

2 Algumas definições

Na seção anterior, argumentei que juntamente com o crescimento e fortalecimento da Linguística de *Corpus* e consequente criação de *corpora* amplos e representativos com uma variedade de dados

⁵ Agradeço a um dos pareceristas que me lembrou da importância de fazer referência às dificuldades de ordem cultural.

de língua em contextos comunicativos naturais, cresceu nas últimas décadas o número de *corpora* construídos via documentação linguística. A Documentação linguística é um campo de atuação da linguística que se ocupa com a criação de registros de línguas em uso, através da criação de acervos digitais que permitem seu acesso e o uso de dados organizados nessas plataformas. Dados coletados via projetos de documentação, no entanto, são bem diferentes do tipo de dado que está presente em *corpora* construídos a partir dos critérios da Linguística de *Corpus*, a começar pelo fato de que, em projetos de documentação, os dados, embora englobem vários tipos de gênero, são coletados para esse fim, já que muitas das línguas objeto de documentação são pouco utilizadas em meio digital, algumas ágrafas, enquanto que, em um *corpus* “prototípico”, os dados que compõem a amostra foram produzidos com objetivos comunicativos variados em contextos reais de uso da língua. Assim, é preciso distinguir o que é um *corpus* produzido conforme os pressupostos da Linguística de *Corpus*.

Como descreve Sanchez, para a Linguística de *Corpus*, o conceito de *corpus* diz respeito a

um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SANCHEZ, 1995, p.8-9)

Dessa forma, vemos que para essa área o conceito de *corpus* vai muito além da definição geral de “conjunto de documentos que servem de base para a descrição ou estudo de um fenômeno” (Dicionário on-line Priberam, 2020). Segundo Aluísio e Almeida (2006), há um conjunto de requisitos que impactarão na validade e confiabilidade de uma pesquisa segundo os pressupostos da Linguística de *Corpus*; assim, para o projeto de um *corpus* computadorizado, é preciso observar seis características: autenticidade, representatividade, balanceamento, amostragem, diversidade e tamanho. Assim, *corpus* ideal para a Linguística de *Corpus*, deve conter:

1. textos autênticos, o que significa que os textos devem ter sido escritos em linguagem natural, não podendo ser textos produzidos com o propósito de serem alvo de pesquisa linguística e devem ser produzidos por falantes nativos;
2. representatividade, deve ser representativo de uma língua ou variedade de língua que se deseja pesquisar. Nesse sentido, ao selecionar os textos, deve-se perguntar: Quais tipos de textos? Quais gêneros textuais? E se de fato representam os usos linguísticos de uma comunidade;
3. balanceamento, o *corpus* deve ter um equilíbrio de gêneros discursivos (informativo, científico, religioso etc.) ou de tipos de textos (artigo, entrevista, dissertação, carta etc.), ou de títulos, ou de autores ou de todos esses itens em conjunto;

4. amostragem proporcional, o tamanho das subamostras (falantes, registros, variedades, etc.) deve ser proporcional à proporção de falantes, registros, variedade, etc.;
5. diversidade, se o *corpus* pretende representar/ ser representante de uma língua, ele deve apresentar diversidade dialetal, de tópicos, de gênero para permitir diferentes tipos de estudo;
6. tamanho adequado ao tipo de pesquisa que se vai realizar e à metodologia a ser adotada na pesquisa.

Em relação às etapas metodológicas, é preciso (i) selecionar os textos a serem compilados, que podem ser provenientes da web ou textos impressos⁶ (e posteriormente digitalizados), converter em formato txt e, na compilação, seguir as normas legais de anonimato e para a obtenção de direitos autorais; (ii) estabelecer os níveis de representação das informações presentes num *corpus*: anotação estrutural e anotação linguística. A anotação estrutural corresponde à marcação de dados externos (metadados textuais) e internos (segmentação – capítulos, parágrafos, etc.) dos textos. Já a anotação linguística trata dos níveis que se deseja representar (morfo sintático, sintático, semântico, discursivo, etc.), que podem ser inseridos manualmente (por linguistas), automaticamente (por ferramentas de PLN) ou semi-automaticamente (correção manual de saída de outras ferramentas).

Considerando todas as características apresentadas de 1 a 6, muitos *corpora* importantes disponíveis não poderiam ser definidos como prototípicos. Como a Linguística de *Corpus* lida com textos produzidos com objetivos comunicativos reais, é preciso que haja textos disponíveis em diferentes gêneros e tipologias e com possibilidade de liberação de direitos autorais. Assim, se o objetivo é construir um *corpus* de uma língua com grande presença digital, é possível selecionar textos de diferentes gêneros e tipologias, já em formato digital e de acesso livre (blogs, postagens em mídias sociais, artigos etc.). Se, ao contrário, a língua objeto tem pouca presença digital, ou se está restrita a apenas alguns contextos comunicativos, os critérios de diversidade, balanceamento e mesmo de amostragem proporcional não são cumpridos. Vale lembrar que, em sociedades multilíngues, há falantes que distinguem os contextos comunicativos em que usam uma ou outra língua, ficando algumas línguas restritas ao contexto doméstico e outras às trocas comerciais. Também nos *corpora* históricos, só é possível encontrar as seis características em *corpora* de línguas com uma longa tradição escrita, o que acaba restringindo os *corpora* a algumas línguas europeias, asiáticas e ao árabe, já que muitas línguas são ágrafas, ou criaram muito recentemente um sistema de escrita, ou surgiram há pouco tempo. Em relação às línguas indígenas brasileiras, por exemplo, há pouquíssimos registros históricos escritos e em apenas alguns poucos gêneros – gramáticas (BATISTA, 2005) catecismos, dicionários

6 As tecnologias de processamento de fala ainda têm dificuldade de lidar com diferentes frequências de voz, dialetos e outros elementos difíceis de parametrizar. Por isso, mesmo os textos orais são antes transcritos.

(Enciclopédia das línguas no Brasil, 2020).

Já o critério de tamanho se relaciona ao critério de representatividade, embora isso seja bastante relativo, uma vez que depende muito dos objetivos de pesquisa. Assim, é possível que um *corpus* seja representativo mesmo sendo de tamanho pequeno. Apenas a título de ilustração, apresento na Tabela 1 uma escala de tamanho proposta por Berber Sardinha (2002, p.119) baseada na análise de trabalhos completos publicados em cinco eventos dedicados à Linguística de *Corpus* (três edições do ICAME – *International Computer Archive of Modern and Medieval English*, uma edição do PALC – *Practical Applications in Language and Computers* e uma edição do TALC – *Teaching and Language Corpora*):

Tabela 1: Classificação de *corpora* de acordo com o número de palavras

Classificação do <i>corpus</i>	Número de palavras
Pequeno	Até 80 mil
Pequeno-médio	80 a 250 mil
Médio	250 mil a 1 milhão
Médio-grande	1 milhão a 10 milhões
Grande	Acima de 10 milhões

Fonte: Berber Sardinha (2002, p.119)

A presença ou não de todas as características de 1 a 6 distingue basicamente dois tipos de *corpora*: (i) os gerais, que pretendem representar a língua de forma ampla e servir de base para pesquisas variadas, caracterizando-se pela variedade de gêneros discursivos, registros, assuntos e autores; e (ii) os especializados, coletados para objetivos específicos de pesquisa e que consistem em coleções de textos de gêneros ou discursos específicos. Como um exemplo de *corpus* geral temos o *British National Corpus* (BNC), criado nos anos 80, que contém 100 milhões de palavras de textos de uma ampla variedade de gêneros já anotados e etiquetados gramaticalmente. Há, claro, *corpora* gerais de várias outras línguas: francês (SIEPMANN; BÜRGE, 2014), espanhol (SANCHEZ *et al.*, 1995), árabe (ARTS *et al.*, 2014), chinês (ZHOU; YOU, 1997), estoniano (HENNOSTE *et al.*, 1998), etc. No Brasil, o desenvolvimento da área da Linguística de *Corpus* é mais recente e muitos dos *corpora* compilados são especializados, como o do Projeto DIRECT (PUC-SP), composto de textos de comunicação no contexto de negócios, e do Projeto NURC, composto de textos de fala culta de diferentes regiões urbanas do país, colhidas em situações pré-estabelecidas (OLIVEIRA, 2016). Dentre os *corpora* gerais, podemos citar o CORPOBRAS - PUC-Rio (OLIVEIRA; DIAS, 2009), que busca ser representativo do português do Brasil, apresentando 27 gêneros discursivos e mais de 1.000.000 de palavras; e o *Corpus Anotado do Português Histórico - Tycho Brahe* (Unicamp), com mais de 3.300.000 palavras.

Em relação a ferramentas computacionais, é importante dizer que todos os *corpora* gerais listados no parágrafo acima têm vários tipos de anotação que permitem a aplicação ou desenvolvimento de diversas ferramentas computacionais de busca. Tomando o COCA – *Corpus of Contemporary American English* como exemplo, por ser um *corpus* bastante amplo, com mais de 1 bilhão de palavras, e balanceado, com oito diferentes gêneros, vemos que o sistema sofisticado de anotação permite vários tipos de busca – por palavras, sintagmas, subsequências, lexemas, morfemas, sinônimos, listas de palavras. Com isso, é possível realizar vários tipos de pesquisas, através da frequência de palavras, sintagmas e construções gramaticais, distinguindo gênero textual e período histórico; é possível, por exemplo, comparar a colocação de palavras mais frequentes em um gênero e outro. Também é possível mapear mudanças em construções sintáticas comparando a frequência e uso dessas construções considerando os textos mais recentes e os mais antigos.

Para línguas minoritárias ou ameaçadas é ainda mais difícil encontrar um *corpus* que cumpra mais do que três dos critérios apontados. É importante lembrar que os *corpora* de línguas ameaçadas são, muitas vezes, oriundos de projetos de documentação, que são sempre menores do que os de línguas de maior *status*, e que não podem, geralmente, ser considerados representativos de um tipo particular de falante, registro e nem mesmo variedade, além de, algumas vezes, serem de acesso restrito aos que participaram de sua criação. É o caso dos projetos de documentação realizados a partir do Programa DOBES que compõem o *The Language Archive* (TLA) localizado no Instituto Max Planck de Psicolinguística em Nijmegen⁷ (<https://dobes.mpi.nl/projects/?lang=pt>). Essa grande base de dados apresenta quatro níveis de acesso: no nível 1, há arquivos que podem ser abertos por qualquer pessoa; no nível 2, há arquivos que podem ser acessados por usuários registrados que assinaram o Código de Conduta. No nível 3, há arquivos disponíveis apenas para usuários aprovados por um contribuinte e, no nível 4, os arquivos são fechados (acessados apenas pelo(s) pesquisador(es) que realizaram a documentação).

No Brasil, há várias iniciativas de documentação linguística e cultural realizadas com as comunidades indígenas, que resultaram em grandes acervos digitais armazenados no Museu do Índio, no Rio de Janeiro, e no Museu Paraense Emílio Goeldi – MPEG (<https://www.museu-goeldi.br>), em Belém do Pará. O MPEG é pioneiro no trabalho de documentação linguística no Brasil, especialmente na documentação das línguas faladas na Amazônia. Segundo Galúcio (2004), o acervo do MPEG possui material de mais de 50 línguas, algumas dessas extremamente ameaçadas. Já o Museu do Índio

⁷ O TLA é especializado no arquivamento e na preservação de gravações linguísticas e de outros tipos de dados, bem como no desenvolvimento de ferramentas linguísticas e de seu arquivamento. Já o programa DOBES (em alemão “*Dokumentation bedrohter Sprachen*”, ou seja “Documentação de línguas ameaçadas”) foi iniciado pela Fundação Volkswagen com o propósito de documentar línguas potencialmente sob perigo de extinção dentro de poucos anos.

possui o maior acervo digital de línguas indígenas do Brasil graças à iniciativa da criação do Programa de Documentação de Línguas e Culturas Indígenas – PROGDOC (<http://progdoc.museudoindio.gov.br>). No entanto, o único *corpus* de língua indígena brasileira de livre acesso e anotado é do kadiwéu vinculado ao Tycho Brahe, na Unicamp (GALVES; SANDALO; SENA; VERONESI, 2017). O projeto kadiwéu é pioneiro ao disponibilizar um *corpus* anotado de língua indígena americana que permite pesquisas gramaticais que incluem informação morfológica e sintática.

De maneira geral, há muitas línguas que não dispõem de qualquer tipo de *corpus* eletrônico. Para se ter uma ideia da pouca amplitude dos *corpora* que temos hoje em relação à quantidade de línguas humanas, basta observar que, embora a maior parte seja de línguas indo-europeias, das 6000 línguas vivas estimadas, 3000, ou seja 50%, são faladas na Ásia, por volta de 1900 na África (31%), 900 nas Américas (15%) e somente 275 são faladas na Europa e no Oriente Médio, o que corresponde a apenas 4%. Ainda mais revelador é que a maior parte dessas línguas está concentrada em 22 países: Papua Nova Guiné (850), Indonésia (670), Nigéria (410), Índia (380), Camarões (270), Austrália (250), México (240), Congo (210) e Brasil (160).

Toda essa discussão em torno da composição da documentação de línguas ameaçadas leva a um outro ponto importante a ser abordado nessa seção; me refiro ao uso do termo *less-resourced languages*. O termo presente no título, como mencionado, tem sua origem na Linguística Computacional e é usado para definir línguas em que falta uma presença digital significativa (SZYMANSKI, 2011).

Eu também poderia ter escolhido os termos ‘línguas pouco estudadas’, ‘ameaçadas’ ou ‘minoritárias’, mas o fato é que, dentro do conceito de *less-resourced languages*, estão incluídas, na maioria das vezes, línguas minoritárias ou ameaçadas, mas não apenas. Há casos de línguas que não podem ser classificadas como ameaçadas, mas que, ainda assim, apresentam pouca presença digital, o que pode ser explicado pela ausência de uma tradição escrita, ou mesmo de uma grafia, e pela falta de acesso a recursos tecnológicos por parte dos falantes. Já o termo ‘línguas pouco estudadas’ é muitas vezes equalizado a *less-resourced languages*, uma vez que na prática a ausência de recursos significa também a ausência de estudos científicos, pelo menos na área de PLN. A escolha entre esses dois termos, deu-se por dois fatores, pela existência de uma referência na área da Linguística Computacional e também porque, neste artigo, advogo justamente pela criação de *corpora*, ainda que não possam ser representativos da totalidade de usos linguísticos de uma língua.

Ainda tratando das definições, Maxwell e Hughes (2006, p.30) preferem o termo *lower density languages*. Segundo os autores, há apenas um pequeno número de línguas do mundo, como o inglês,

o chinês, o árabe e línguas da Europa Ocidental, cujos recursos são abundantes e que poderiam, por isso, serem definidas como línguas de alta densidade. Para algumas outras poucas, os recursos, se não abundantes, pelo menos existem e crescem cada vez mais; essas seriam as línguas denominadas como de densidade média, que é o caso das demais línguas europeias. Ainda segundo os autores, esses dois conjuntos somados não chegam a representar 30% das línguas do mundo. Isso significa que, para a grande maioria das línguas, os recursos são escassos.

Se, como afirmamos acima, a ausência de recursos implica a ausência de estudos científicos, isso significa que a maioria dos estudos linguísticos empíricos está disponível apenas para 30% das línguas humanas. Se faltam recursos para o desenvolvimento de estudos empíricos, conseqüentemente, boa parte de nosso conhecimento científico linguístico fica restrito a um número muito pequeno de línguas. Tal fato leva a crer que, por um lado, a criação de *corpora* poderia ampliar consideravelmente as possibilidades de análise linguística; por outro, lembramos que a amplitude de dados proposta pelos teóricos da Linguística de *Corpus* tem a ambição de possibilitar generalizações do *corpus* para a língua, o que muitas vezes não é possível para línguas com poucos recursos. Assim, qual seria a finalidade da composição de *corpora* para essas línguas? Para responder a essa questão, apresento, nas seções seguintes, algumas discussões sobre a constituição de *corpora* de línguas com poucos recursos.

3 Sobre *corpora* de línguas com poucos recursos

Como mencionado, têm surgido nas últimas décadas vários *corpora* de línguas ameaçadas. O crescimento deste tipo de *corpus* está relacionado à reação por parte da comunidade acadêmica à eminência de extinção de boa parte das línguas humanas. Para se ter uma ideia da magnitude do problema, mesmo as estimativas mais otimistas projetam a perda de 50% das línguas do mundo no século XXI (KRAUS, 1992). Assim, surgiram, desde a década de 1990, vários programas de documentação de línguas ameaçadas, por exemplo, o *Dokumentation Bedrohter Sprachen* (DOBES), financiado pela Fundação *Volkswagem* na Alemanha (2000-2013), e o *Endangered Languages Documentation Programme* – ELDP (<https://www.eldp.net>), financiado pela *US National Science Foundation* (NSF) e pelo *National Endowment of the Humanities* (2005-).

Todo esse processo fez com que a Documentação Linguística, historicamente relacionada à criação de gramáticas, dicionários e coleção de textos, alcançasse o *status* de uma subárea da Linguística. Com isso, os métodos para gravação e análise de materiais linguísticos e culturais foram aprimorados, incluindo-se a criação de bancos de metadados e o uso de tecnologias computacionais.

Um dos pontos de divergência da metodologia de Documentação Linguística em relação à Linguística de *Corpus* é que, enquanto na documentação os dados são, muitas vezes, produzidos especialmente para serem incluídos no *corpus*, para os propósitos da Linguística de *Corpus* é essencial que os dados sejam reais, produzidos com objetivos comunicativos próprios da situação comunicativa real.

Assim, vemos que, embora a composição dos *corpora* coletados através de projetos de documentação seja diferente daquela dos *corpora* formados a partir da metodologia da Linguística de *Corpus*, é relevante dizer que eles existem e são usados na descrição e análise linguísticas. Já para um grande número de línguas não há qualquer tipo de *corpus*, isso porque somente 1% das línguas do mundo tem a vantagem de serem reconhecidas como línguas oficiais, apesar de seus falantes chegarem a compor 60% da população mundial. Além da população de falantes, o *status* dessas línguas também é desigual, o que também tem impacto sobre como *corpora* são criados e usados. McEnery e Ostler (2000, p. 405) apontam pelo menos oito classes de línguas que têm problemas para encontrar apoio governamental para realizar a construção de *corpus*:

- (1) Línguas regionais oficiais sem status nacional (e.g. galês);
- (2) Dialeto – variantes de uma ampla língua nacional (e.g. variedades do chinês);
- (3) Vernáculos – línguas com um grande número de falantes sem reconhecimento oficial (e.g. Sylhet em Bangladesh) ;
- (4) Línguas nacionais com uma população pequena de falantes (e.g. islandês);
- (5) Línguas minoritárias – vernáculos com um número relativamente pequeno de falantes (e.g. a língua gaélica da Ilha de Man, no Reino Unido – manês);
- (6) Línguas minoritárias não-indígenas, é o caso das línguas de imigrantes (e.g. Punjabi, no Reino Unido)
- (7) Línguas ameaçadas (e.g. Kereque, com 100 falantes);
- (8) Línguas de sinais (e.g. LIBRAS).

A esta lista, eu acrescentaria (9) Línguas extintas (e.g. latim, tupi), uma vez que há muitas línguas já extintas que apresentam, porém, textos escritos que poderiam compor um *corpus*. Nesse caso, também há particularidades, já que é praticamente impossível encontrar *corpora* históricos amplos e balanceados.

É importante retomar que a característica fundamental para a definição de línguas com poucos recursos é a ausência de recursos computacionais disponíveis, muito mais do que número de falantes ou quantidade de pesquisas realizadas. Assim, a dificuldade fundamental para o uso mais amplo de

corpora de línguas ameaçadas ou minoritárias está nos problemas com a anotação, que é ausente em alguns *corpora* ou deficiente em outros. Aqui é importante mencionar que a anotação é essencial para a pesquisa linguística envolvendo *corpora*, uma vez que *corpora* anotados podem ser usados para treinar algoritmos através de aprendizagem de máquina, o que permite a criação de recursos computacionais.

As possibilidades de anotação e geração de recursos é bastante ampla, há (i) texto paralelo alinhado com outra língua no nível sentencial e, às vezes, com um ou mais níveis de paralelismo; (ii) texto anotado por entidade nomeada em vários níveis de granularidade; (iii) textos analisados morfológicamente, também com um esquema de rótulos morfológicos apropriados para uma língua em particular; (iv) textos marcados por fronteiras de palavras, importante especialmente para línguas que não marcam a maioria das fronteiras de palavras; (v) texto com rótulo POS (*part of speech*) e com um esquema de rótulos POS apropriados para a língua particular (e.g. N (*noun* - substantivo), V (*verb* - verbo), P (*preposition* - preposição etc.); (vi) banco de árvores (sintaticamente anotadas e parseadas); (vii) texto rotulado semanticamente e (viii) dicionários eletrônicos e outros recursos lexicais, como Wordnet⁸.

Voltando às diferenças para a composição de bancos de dados entre a Linguística de *Corpus* e a Documentação Linguística, é relevante dizer que muitos *corpora* de documentação são de acesso restrito aos pesquisadores que atuaram em sua criação; o que torna impossível apresentar nessa seção uma análise completa da organização e funcionalidade de *corpora* de dessas línguas. Por esse motivo, minha argumentação parte de um tipo específico de *corpora* de línguas com poucos recursos, me refiro a *corpora* de tamanhos entre médio e grande de línguas conhecidas, mas com baixa presença digital. Assim, selecionei dois *corpora* sem qualquer tipo de anotação e dois anotados⁹ para dar ao leitor uma ideia geral da diferença desses *corpora* em relação aos prototípicos.

Dentre *corpora* não-anotados, Vinogradov (2016) cita o Assamese e o Ndebele. O assamese, segundo o autor, é uma língua indo-iraniana falada por quase 13 milhões de pessoas na Índia. O *corpus* do assamese é composto de 3 milhões de *tokens* (palavras, números etc) de 1191 textos ao todo e está integrado ao Projeto EMILLE – Enabling Minority Language Engineering. Os textos são de um conjunto bastante variado de gêneros, mas, como não é anotado, a única forma de pesquisa no *corpus* é através da forma exata em que as palavras estão escritas. Já o ndebele é uma língua africana vinculada ao grupo banto da macro-família Níger-Congo e falada por aproximadamente um milhão

8 <http://wordnet.princeton.edu>

9 Consideramos como anotado mesmo aquele corpus em que há apenas um conjunto de dados anotados.

e meio de pessoas. O *corpus* dessa língua, composto por textos escritos e falados, contém 691.268 *tokens* e foi desenvolvido como parte do Projeto ALLEX – African Languages Lexicon. Esse último, no entanto, é voltado para a produção de dicionários e outras ferramentas para línguas africanas usadas no Zimbabué, o que o caracteriza como um *corpus* especializado.

Os dois *corpora* citados acima são de tamanho, respectivamente, médio-grande e médio, mas enquanto o *corpus* do assamese apresenta uma grande variedade de gêneros, o do ndebele distingue apenas entre textos escritos e falados, sem mencionar gênero. Nesse caso são dois *corpora* de bom tamanho e um deles bastante diverso, mas não contam com nenhum sistema de anotação.

Para seguir o mesmo padrão, selecionei também para a descrição dos *corpora* anotados uma língua iraniana, o osseto, e uma africana, o bambara. O osseto é falado por aproximadamente meio milhão de pessoas na parte central do Cáucaso (Federação Russa e Geórgia). O *corpus* do osseto é composto de textos escritos e contém mais de 11 milhões de *tokens*. Como é anotado, inclui informações gramaticais sobre os *tokens* e tradução. Segundo Vinogradov (2016), o principal mérito desse *corpus* está no mecanismo de busca e na interface amigável. Nesse *corpus*, é possível fazer buscas por lexema, forma da palavra, tradução ou por um conjunto de traços gramaticais. Também é possível selecionar mais de um *token* e checar a distância entre eles.

O bambara é uma língua Mandê que, como o ndebele, pertence à macro-família Níger-Congo. Esse é um dos poucos casos de língua com poucos recursos, mas não ameaçada, já que é falada por mais de 10 milhões de pessoas. O *corpus* do bambara é composto de 426.813 *tokens* de textos de diferentes gêneros e zonas dialetais. Como também é anotado e contém subcorpora desambiguados, permite ao usuário a pesquisa por lexema, forma das palavras, sintagmas, símbolos, classe de palavras, além de especificar o contexto e permitir a visualização dos resultados em diagramas de frequência.

A partir da descrição desses quatro *corpora*, é possível observar que as únicas duas características comuns entre eles são a legibilidade por máquina e a finitude, já que esses bancos de dados diferem quanto à quantidade e diversidade dos textos disponíveis para cada língua em particular. Em alguns casos, há uma drástica diferença na distribuição dos textos de acordo com o gênero, em outros, os dados não são autênticos, tendo sido produzidos via elicitación. Assim, vemos que não há uma uniformidade de tipos de dados e recursos entre os *corpora* dessas línguas, uma vez que a motivação para a sua criação não é sempre a mesma e eles tendem a ser mais especializados do que gerais.

Há ainda outro ponto que chama a atenção, são as diferenças em relação à possibilidade de uso de ferramentas computacionais. Os *corpora* não anotados, é claro, só permitem pesquisas

simples através da forma das palavras, mas mesmo entre os dois *corpora* anotados descritos acima, há diferenças quanto ao tipo de anotação realizado. Em alguns casos, o padrão de anotação permite uma grande variedade de buscas, o que o torna uma poderosa ferramenta para a descrição e análise linguísticas, mas ainda assim, as possibilidades de descrição e análise são bem distantes do que está disponível para os *corpora* de línguas com ampla presença digital. Relevante nesse caso é a quantidade de textos escritos produzidos. Se uma língua é ágrafa, mesmo sendo falada por um grande número de indivíduos, isso significa certamente que haverá pouca presença digital.

As dificuldades enfrentadas na documentação e criação de *corpora* de línguas ágrafas são tão reconhecidas que houve recentemente a criação de um projeto voltado para esse fim. O *Breaking the Unwritten Language Barrier* (BULB), como é chamado, aproxima linguistas e cientistas computacionais com o objetivo de desenvolver ferramentas que sirvam de auxílio para o trabalho de documentação de línguas ágrafas, como reconhecimento automático de fala e tradução por máquina (ADDA; ADDA-DECKER; AMBOUROUE; BESACIER; BLACHON *et al*, 2016).

Há ainda outras razões para a baixa frequência de textos eletrônicos (McENERY; OSTLER, 2000, p. 411): (1) baixo nível de letramento da comunidade; (2) um sistema de escrita ainda não computadorizado; e (3) um sistema de escrita computadorizado, mas sujeito à competição com padrões.

Em relação aos problemas de padronização, o Unicode¹⁰, que se refere a um conjunto universal de caracteres para o sistema de escrita humana, é apontado como uma esperança para se resolver a questão, embora não seja capaz de sozinho resolver todos os problemas para a composição de *corpora*.

Há ainda outros problemas de ordem tecnológica enfrentados para a constituição de *corpora* eletrônicos para línguas com poucos recursos. As ferramentas e técnicas de processamento de texto existentes foram em sua maioria desenvolvidas a partir de aprendizagem de máquina, que precisa de uma grande quantidade de dados. Na falta desses, muitos programadores optam por sistemas de regras, o que, na verdade, pouco auxilia o trabalho do linguista que vai trabalhar com línguas pouco estudadas, já que para se criar ferramentas via sistema de regras é necessário já conhecer de antemão o funcionamento da língua.

Resumindo, quando se compara os processos de composição de *corpora* de línguas com poucos recursos com os considerados prototípicos, observa-se que os primeiros são mais limitados, alguns

10 Veja <http://www.unicode.org/unicode/standard/standard.html> para mais detalhes.

não são bem distribuídos em termos de gêneros e variedades, outros são bem distribuídos, mas não são anotados. Assim, vale refletir sobre que contribuições a constituição de bancos de dados de línguas com poucos recursos podem trazer, já que muitas vezes não são amplos o suficiente para serem representativos de toda a língua.

4 Por que criar corpora para línguas com poucos recursos?

Dadas as peculiaridades na composição de *corpora* de línguas com poucos recursos descritos na seção anterior, torna-se inevitável questionar se eles realmente valem a pena e que contribuições podem trazer. Para começar a reflexão, é interessante observar a seguinte afirmação de McEnery e Ostler (2000, p.403), “se a linguística de *corpus* é uma abordagem útil na linguística, então ela deveria ser aplicada a todas as línguas”. A partir desse imperativo, vale pensar que a Linguística de *Corpus* é uma área multidisciplinar de interesse não só de linguistas, mas também de cientistas computacionais, tecnólogos. E que há pelos menos dois usos possíveis para um *corpus*, como de fonte de dados e como *testbed* para sistemas. Isso significa que quanto mais *corpora*, maiores são as fontes de dados para análises linguísticas e para profissionais de PLN. Para o linguista, o acesso a dados de línguas pouco estudadas pode significar mais argumentos a favor de uma hipótese ou colocar na berlinda teorias que não dão conta de explicar fenômenos raros. Para a área de PLN, o acesso a esses dados pode se apresentar como um desafio e um motivo para a criação de novas ferramentas de processamento, já que, nesse caso, o profissional não poderá sempre usar tecnologias já existentes, pensadas a partir de dados de algumas poucas línguas.

Um outro argumento é que a criação de *corpora* permite que fiquem armazenados conhecimentos de todo tipo, desde históricos e procedimentais a processos evolutivos das sociedades. Também nesse sentido, vale lembrar que a maior parte dos *corpora* de línguas com pouco recursos são oriundos de projetos de documentação de línguas ameaçadas, o que significa que podem ser o único registro dessas línguas, o que é de valor inestimável.

E ainda em relação à documentação de línguas ameaçadas, é importante lembrar que dependendo do grau de ameaça, o tempo que o linguista terá para documentar o máximo possível de dados da língua é muito curto, especialmente quando não há descrições sobre a língua. Se houvesse mais ferramentas computacionais que automatizassem a anotação e a tradução dos dados, o linguista ganharia mais tempo para realizar a documentação e interpretação dos dados coletados. Vale salientar que a pesquisa de campo, para documentação linguística, além de demorada, é cara, e o desenvolvimento de tecnologias computacionais além de otimizar o trabalho, também diminuiria os custos desse tipo de trabalho.

As possibilidades de aplicação de tecnologia linguística são ainda um fator a ser considerado. Como comentam McEnery e Ostler (2000), a maior parte dos sistemas de processamento de língua foram desenvolvidos a partir de línguas de comunidades ricas e poderosas (Estados Unidos, Europa, Japão), enquanto os *corpora* de comunidades pobres e pequenas foram designados simplesmente como repositórios culturais. Uma mudança nesse aspecto poderia, como argumentam os autores, aumentar o acesso de comunidades de línguas não-majoritárias a essas tecnologias e ampliar seu conhecimento técnico e letramento nas novas mídias.

Um último ponto que gostaria de destacar é que para além das possibilidades de desenvolvimento das áreas da Linguística e da Ciência da Computação, a constituição de *corpora* pode ser muito relevante também para as comunidades envolvidas, já que a produção de recursos e o interesse pela língua tendem a resultar na promoção desta entre os membros da comunidade. Como já bastante discutido, a melhora no *status* de uma língua é um dos pontos fundamentais para promovê-la e revitalizá-la.

Por fim, diante de todos os argumentos levantados nessa seção, entendo que a constituição de *corpora* para línguas com poucos recursos, ainda que não cumpram todos os critérios propostos pela Linguística de *Corpus*, devem ser realizados, e os resultados devem ser aproveitados de diversas formas, seja gerando novas tecnologias, servindo de suporte empírico para teorias linguísticas ou promovendo a língua na comunidade.

Considerações Finais

Iniciamos esse artigo falando do desenvolvimento da área da Linguística de *Corpus*, apresentamos as características de *corpora* “prototípicos” e discutimos as dificuldades para se criar *corpus* gerais que possam ser representativos dos usos de toda uma língua ou variedade, especialmente para *corpora* históricos e de línguas com pouca presença digital. Como vimos, os teóricos da área entendem que um *corpus* deve ser amplo, balanceado e diverso, cobrindo a maior variedade possível de usos de uma língua e apresentando uma boa distribuição entre as subamostras, além de terem de ser finitos e em formato que possa ser processado pelo computador. Seriam, no entanto, poucas as línguas com recursos suficientes para resultarem em um *corpus* balanceado e realmente representativo de todos os usos de uma língua.

A partir de dados sobre a distribuição das línguas no mundo, vimos que menos de 30% delas contam com recursos suficientes para gerarem um *corpus* prototípico e, conseqüentemente, apenas 30% delas podem contar com estudos linguísticos empíricos. E mais ainda, vimos que a maioria

das teorias linguísticas vigentes são resultado do nosso conhecimento de menos de 4% das línguas humanas, já que apenas línguas oficiais de grandes estados-nação contam com apoio governamental para a sua documentação.

Também mencionei o fato de que tem surgido nas últimas décadas uma série de *corpora* menores, coletados em projetos de documentação linguística dada a iminência de extinção da maior parte das línguas no mundo já neste século. Como discutido, esses *corpora* têm características diferentes dos *corpora* prototípicos, o que torna difícil o desenvolvimento de recursos tecnológicos.

Ao final, concluo que mesmo que os *corpora* de línguas com poucos recursos não tenham a mesma amplitude e os mesmos recursos dos ditos prototípicos, eles são extremamente importantes por várias razões. Podem sim ser uma excelente fonte de dados linguísticos, podem permitir o avanço de tecnologias de PLN, ampliar o acesso a tecnologia e ainda promover o *status* de uma língua entre a comunidade.

REFERÊNCIAS

ADDA, Gilles; ADDA-DECKER, Martine; AMBOUROUE, Odette; BESACIER, Laurent; BLACHON, David *et al.* Innovative technologies for under-resourced language documentation: The BULB Project. In: WORKSHOP CCURL 2016 – COLLABORATION AND COMPUTING FOR UNDER-RESOURCED LANGUAGES – LREC, 2016, Eslovênia, *Proceedings* [...] Eslovênia, HAL, 2016. Disponível em: <https://hal.archives-ouvertes.fr/hal-01350124>. Acesso em: 10 jan. 2020.

ALLEX – African Languages Lexicon. 2006. Disponível em: <http://www.edd.uio.no/allex/> Acesso em: 10 jan. 2020.

ALUISIO, Sandra M.; ALMEIDA, Gladis M. B. O que é e como se constrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. *Caleidoscópio*, v. 4, n. 3, p. 156-178, 2006.

ARTS, Tressy *et al.* arTenTen: Arabic Corpus and Word Sketches. *Journal of King Saud University – Computer and Information Sciences*. v. 26, n. 4, p. 357-371, 2014.

ASTON, Guy; BURNARD, Lou. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edimburgo: Edinburgh University press, 1998.

BATISTA, Ronaldo de O. Descrição de Línguas indígenas em gramáticas missionárias do Brasil

Colonial. *D.E.L.T.A*, v. 21, n. 1, p. 121-147, 2005.

BERBER SARDINHA, Tony. *Linguística de corpus*. São Paulo: Manole, 2004.

_____. Tamanho de corpus. *The ESP*, São Paulo, v. 23, n. 2, p. 103-122, 2002.

_____. Linguística de Corpus: Histórico e Problemática. *Delta*, São Paulo, v.12, n.2, p. 323-367, 2000.

BRITISH NATIONAL CORPUS (BNC). Disponível em: <<https://www.english-corpora.org/bnc/>>. Acesso em: 03 jan. 2020.

COCA – Corpus of Contemporary American English. Disponível em: <<https://www.english-corpora.org/coca/>>. Acesso em: 04 jan. 2020.

COX, Christopher. Corpus Linguistics and Language Documentation: Challenges for Collaboration. In: NEWMAN, J.; BAAYEN, H.; RICE, S. (eds.) *Corpus-Based Studies in Language Use, Language Learning, and Language Documentation*. Amsterdam: Rodopi, 2011. p. 239-264.

DIRECT. 2005. Disponível em: <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>. Acesso em: 03 jan. 2020.

DOBES. 2006. Disponível em: <<https://dobes.mpi.nl/projects/?lang=pt>>. Acesso em: 02 dez. 2019.

ELDP. 2005. Disponível em: <https://www.eldp.net>. Acesso em: 02 dez. 2019.

EMILLE – Enabling Minority Language Engineering. 2003. Disponível em: <https://www.lancaster.ac.uk/fass/projects/corpus/emille/>. Acesso em: 10 jan. 2020.

ENCICLOPÉDIA DAS LÍNGUAS NO BRASIL. Disponível em: https://www.labeurb.unicamp.br/elb/indigenas/documentacao_linguas_indigenas.html. Acesso em: 13 jan. 2020.

GALVES, Charlotte C.; SANDALO, Filomena; SENA, Ticiania A. de; VERONESI, Luis. Annotating a polysynthetic language: From Portuguese to Kadiwéu. *Cadernos de Estudos Linguísticos*. Campinas, p. 631-648, set./dez. 2017.

GALÚCIO, Ana Vilacy. Gravações e acervo a partir da pesquisa linguística e cultural, como um passo para a revitalização, fortalecimento e resgate cultural. In: MOREIRA, E; BELAS, C.A.;

et al (orgs.). *Propriedade Intelectual e Patrimônio Cultural: proteção do conhecimento e das expressões culturais tradicionais*, Belém do Pará, MPEG, 2004. p. 109-115.

GRIES, Stefan Th. What is Corpus Linguistics? *Language and Linguistics Compass*, v. 3, p. 1-17, 2009.

GRIES, Stefan Th.; BEREZ, Andrea L. Linguistic Annotation in/for Corpus Linguistics. In: IDE, N.; PUTEJOVSKY, J. *Handbook of Linguistic Annotation*. Berlin – New York: Springer, 2015.

HENNOSTE, Tiit *et al.* Structure and usage of the Tartu University Corpus of Written Estonian. *International Journal of Corpus Linguistics*, v. 3, n.2, p. 279-304, 1998.

JOKINEN, Kristiina. Researching Less-Resourced Languages – the DigiSami Corpus. In: ELEVENTH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC). *Proceedings [...]* 2018. p. 3382-3386.

KRAUSS, Michael. The World's Languages in Crisis. *Language*, v. 68, p. 4-10, 1992.

LÜDELING, Anke; KYTÖ, Merja. *Corpus Linguistics: An International Handbook*. V. 1. Berlin – New York: Walter de Gruyter, 2008.

MAXWELL, Mike; HUGHES, Baden. Frontiers in Linguistic Annotation for Lower-Density Languages. In: WORKSHOP ON FRONTIERS IN LINGUISTICALLY ANNOTATED CORPORA. Sydney. *Proceedings [...]*. Sydney, 2006. p. 29-37.

McENERY, Tony; WILSON, Andrew. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press, 2001.

McENERY, Tony; OSTLER, Nicholas. A New Agenda for Corpus Linguistics – Working with all of the World's Languages. *Literary and Linguistic Computing*, v. 15, n. 4, p. 403-419, 2000.

OLIVEIRA, Lúcia P.; DIAS, Maria Carmelita P. Compilação de corpus: representatividade e o CORPOBRAS. *Calidoscópico*. v. 7, n. 3, 192-198, set/dez, 2009.

OLIVEIRA Jr., Miguel. NURC Digital. Um protocolo para a digitalização, anotação, arquivamento e disseminação do material do Projeto da Norma Urbana Linguística Culta (NURC). *CHIMERA. Romance Corpora and Linguistic Studies*, v. 3, n. 2, p. 149-174. 2016.

OSTLER, Nicholas. Corpora of less studied languages. In.: LÜDELING, A.; KYTÖ, M. *Corpus Linguistics: An International Handbook*. Volume 1. Berlin – New York: Walter de Gruyter, 2008. p. 457-483

SANCHEZ, Aquilino. Definición e historia de los corpus. In: A. SANCHEZ et al (org.). *CUMBRE – Corpus Lingüístico de Español Contemporáneo*. Madrid: SGEL, 1995. p. 7-24.

SCANNELL, Kevin P. The Crúbadán Project: Corpus Building for Under-Resourced Languages. *Cahiers du Central*, v. 5, n. 1, p. 1-10, 2007.

SIEPMANN, Dirk; BÜRGEL, Christoph. Le corpus de référence du français contemporain (CRFC), 2014. Disponível em: <https://zenodo.org/record/12353?ln=en#.Xqd1Uy_OpZp>. Acesso em: 27 abr. 2020.

SINCLAIR, Jane M. Corpus and text. Basic principles. In: WYNNE (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 2005. p. 1-16.

SZYMANSKI, Terrence D. *Morphological Inference from Bixtext for Resource-Poor Languages*. 2011. Tese (Doutorado em Linguística) – Universidade de Michigan, Michigan, 2011.

TYCHO BRAHE. 2008. Disponível em <<http://www.tycho.iel.unicamp.br/~tycho/pesquisa/>>. Acesso em 10 dez. 2019.

TRASK, Robert L. *Dicionário de Linguagem e Linguística*. São Paulo: Contexto, 2004.

VINOGRADOV, Igor. Linguistic corpora of understudied languages: Do they make sense? *Káñina, Revista Artes y Letras, Universidad Costa Rica XL*, v.1, p. 127-141, 2016.

ZHOU, Qiang; YU, Shiwen. Annotating the Contemporary Chinese Corpus. *International Journal of Corpus Linguistics*, v. 2, n. 2, p. 199-238. 1997.

**AVALIANDO A VITALIDADE LINGUÍSTICA EM CONTEXTOS DE
MULTILINGUISMO: ETNOGRAFIAS VERSUS MODELOS COMPUTACIONAIS**
*ASSESSING LANGUAGE VITALITY IN MULTILINGUAL CONTEXTS:
ETHNOGRAPHIES VERSUS COMPUTACIONAL MODELS*

Evani Viotti¹

RESUMO

O objetivo deste trabalho é o de discutir o alcance de estudos sobre a vitalidade linguística em regiões multilíngues, feitos no âmbito da teoria da complexidade por meio de modelagens computacionais. Um desses estudos será descrito e comparado com relatos etnográficos que expõem a dinâmica das relações socioeconômicas e políticas que caracterizam algumas ecologias multilíngues, como a do Alto Rio Negro na Amazônia brasileira. Considerando que assumir a língua como um sistema complexo, aberto, dinâmico e auto-organizado é uma avenida promissora para as investigações linguísticas, este trabalho sugere que etnografias e descrições sociolinguísticas parecem capturar mais adequadamente a dinamicidade e complexidade da língua como um fenômeno social do que alguns modelos computacionais que precisam simplificar sobremaneira a caracterização do sistema, e que se baseiam antes em escolhas feitas pelos pesquisadores do que em fatores que, de fato, são definidores do sistema. **Palavras-chave:** etnografias; modelos computacionais; vitalidade linguística; Alto Rio Negro

ABSTRACT

This paper aims at discussing the import of studies on language vitality in multilingual ecologies, carried out within the framework of complexity science by means of computer models. One of these studies will be described and compared with ethnographic reports which illustrate the dynamics of the socioeconomic and political relations which characterize some multilingual ecologies, in particular the Upper Rio Negro in Brazilian Amazonia. Considering that assuming language as a complex, open, dynamic and self-organized system is a promising path for linguistic investigations, this work suggests that ethnographies and sociolinguistic descriptions seem to capture the dynamicity and complexity of language as a social phenomenon better than computational models which simplify the characterization of the system, and which are based on factors chosen by the researchers a priori, rather than factors which actually define the system. **Keywords:** ethnographies; computer models; language vitality; Upper Rio Negro

¹ Departamento de Linguística, Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. E-mail: viotti@usp.br.

Agradeço meus colaboradores em São Gabriel da Cachoeira: Américo Sokot e família, dos hupd'äh, e Marlene Inambu, dos baré. Agradeço, ainda, à FAPESP pelo suporte financeiro à pesquisa que embasa este trabalho por meio dos processos 2014/50764-0 e 2018/14359-4. Agradeço também aos pareceristas anônimos, pela leitura atenta do trabalho e pelas valiosas sugestões. Todas as inadequações que talvez ainda permaneçam são de minha inteira responsabilidade.

Introdução

Uma das maiores preocupações da linguística contemporânea tem sido a vitalidade das línguas, em particular daquelas consideradas minoritárias e ainda pouco descritas. Embora seja sabido que todas as línguas vivem em constante transformação e que a extinção de algumas delas – especialmente em contextos de grande contato linguístico – seja um fenômeno natural e largamente atestado pelos estudos de evolução das línguas (MUFWENE, 2003; 2004; 2017), a inquietação pela possibilidade de perda de algumas línguas se mantém grande (HALE *et al.*, 1992).

Tradicionalmente, estudos sobre as ecologias de prática linguística em que a vulnerabilidade de certas línguas tem sempre sido posta em relevo são feitos no âmbito das interfaces entre a linguística e a antropologia (e.g. a linguística antropológica) e entre a linguística e a sociologia (e.g. a sociolinguística). Mais recentemente, novas perspectivas têm se aberto para acessar o fenômeno da vitalidade linguística nas esferas de estudos que tomam a língua como um sistema complexo, dinâmico e adaptativo. Dentro dessas perspectivas, modelos computacionais têm sido utilizados para levantar as possibilidades de escolhas de uma ou algumas línguas em detrimento de outras, feitas pelos falantes que vivem em contextos em que mais de uma língua é usada. Loureiro-Porto e San Miguel (2017) é um exemplo de estudo computacional que aborda os fatores que desempenham algum papel nas escolhas linguísticas dos falantes em ecologias multilíngues e que possam ter impacto na sobrevivência ou perda de uma ou algumas línguas que compõem a ecologia.

Este trabalho tem natureza teórico-metodológica. Seu objetivo é o de avaliar a pertinência descritiva e o alcance de estudos como o de Loureiro-Porto e San Miguel, cotejando-o com relatos etnográficos que expõem as relações sociais entre povos que compõem ecologias multilíngues em que uma ou mais línguas estão em perigo de extinção, como é o caso do Alto Rio Negro na Amazônia brasileira. Por sua natureza, o trabalho não traz uma análise de dados, nem tem como meta fazer uma revisão completa e ampla da literatura pertinente. Seu foco está na problematização de duas diferentes abordagens teórico-metodológicas, que têm como objetivo comum a investigação dos fatores que podem levar à manutenção, mudança ou perda linguística em contextos caracterizados como multilíngues. A seção seguinte traz uma breve descrição da ecologia linguística do Alto Rio Negro. Na sequência, há uma apresentação sucinta das bases das teorias que tratam a língua como um sistema complexo, dinâmico e adaptativo para contextualizar a apresentação que será feita do estudo de Loureiro-Porto e San Miguel, e circunstanciar as questões colocadas por este trabalho. Por fim, os resultados obtidos pelo modelo computacional desses autores serão discutidos à luz de algumas descrições etnográficas de ecologias multilíngues, para mais bem configurar a vitalidade linguística

na região do Alto Rio Negro.

2 O Alto Rio Negro

Povos ancestrais habitam a porção noroeste da bacia amazônica conhecida como ‘Cabeça do Cachorro’, na fronteira do Brasil com a Colômbia e com a Venezuela. Grupos étnicos que constituem a família linguística naduhup (EPPS, 2007; EPPS; BOLAÑOS, 2017; EPPS, 2018) vivem na região há pelos menos 600 anos, talvez mais (NEVES, 1998).² Grupos da família arawak vindos do norte (NIMUENDAJÚ, 1950; NEVES, 1998), e, mais tarde, povos tukano oriental – possivelmente vindos do oeste – juntaram-se a eles,³ constituindo parte do que veio a se tornar um dos mais ricos sistemas multilíngues da América do Sul: um pouco mais de 20 línguas diferentes são faladas na região (EPPS; STENZEL, 2013).

A chegada dos povos arawak e tukano nessa porção da Amazônia parece ter sido o fator que levou alguns dos povos da família naduhup a ocupar as zonas interfluviais entre os rios Tiquié e Papuri,⁴ afluentes da margem esquerda do rio Vaupés (por sua vez, afluente do Rio Negro). Eles se caracterizam por originalmente ser povos seminômades que preferem se locomover a pé (não por barcos), e são fundamentalmente caçadores e coletores, praticando apenas uma agricultura de subsistência. Por esses motivos, eles são conhecidos como *Povos da Floresta*. Diferentemente, os povos das famílias linguísticas arawak e tukano oriental vivem nas proximidades dos grandes rios, locomovem-se preferencialmente por barcos, têm a pesca como uma de suas atividades centrais, e têm uma prática de agricultura mais variada e desenvolvida. Eles são conhecidos como *Povos do Rio* (EPPS; STENZEL, 2013; EPPS; BOLAÑOS, 2017; EPPS, 2018).

Apesar das claras distinções que separam Povos da Floresta e Povos do Rio em termos de suas línguas, práticas matrimoniais e padrões de subsistência (EPPS; BOLAÑOS, 2017), eles interagem de várias maneiras. Por vezes, seus encontros são breves, limitados à troca de mercadorias. Povos da Floresta aceitam trabalhar nas roças ou casas dos Povos do Rio, em troca de produtos agrícolas e mercadorias não indígenas, ou pela oportunidade de participar dos *dabucuris*, festivais rituais que

2 Embora a literatura sobre povos e línguas indígenas brasileiras opte, em geral, por grafar a primeira letra dos nomes de etnias e línguas indígenas em maiúsculo, eu preferi seguir o padrão da língua portuguesa, segundo o qual gentílicos e línguas se escrevem com letras minúsculas.

3 As histórias dos povos tukano oriental relatando sua origem sugerem, diferentemente, que eles vieram do leste (cf., por exemplo, MAIA; MAIA, 2004).

4 Refiro-me, em particular, aos povos hupd’äh e yuhupd’äh. Participam também da família naduhup os povos falantes das línguas dâw e nadëb, que vivem mais próximos, respectivamente, das cidades de São Gabriel da Cachoeira e de Santa Isabel do Rio Negro. Para evidências linguísticas que justificam a categorização das línguas hup, yuhup, dâw e nadëb como uma família, ver Epps e Bolaños (2017) e Epps (2018).

podem chegar a durar alguns dias; nessas festas, são contadas histórias ancestrais e os membros dos dois grupos compartilham *caxiri* (bebida fermentada) e tabaco, e participam das rodas de coca (JACKSON, 2012). Em outras ocasiões, o relacionamento entre eles pode se tornar longo, chegando, por vezes, a atravessar gerações. Nesses casos as relações entre os dois grupos de povos se tornam mais complexas, sendo, por vezes, caracterizadas como um sistema assimétrico em que os Povos do Rio gozam de uma posição de privilegiada em relação aos da Floresta (cf. JACKSON, 2012; EPPS; STENZEL, 2013). Os Povos do Rio tendem a chamar os Povos da Floresta pela palavra *maku*, um termo de origem arawak que significa ‘o que não fala’ (ATHIAS, 1995); consideram a língua deles como tendo características animais e impossível de aprender; e tomam-nos como infantis, desorganizados, irresponsáveis e incestuosos, na medida em que eles seguem um sistema de casamento baseado na endogamia linguística (JACKSON, 2012; EPPS; STENZEL, 2013; EPPS; BOLAÑOS, 2017).

Em resumo, o Alto Rio Negro se configura como um quadro de equilíbrio tenso: enquanto algumas práticas, como a língua, o sistema matrimonial e as atividades de subsistência, se mantêm claramente distintas, outras se mostram bastante congruentes. A divisão entre Povos do Rio e Povos da Floresta, que à primeira vista parece revelar uma complementaridade socioeconômica saudável entre habitantes de uma vasta região da Amazônia brasileira, na verdade envolve uma delicada e difícil configuração de poder, em que os Povos do Rio se consideram como em uma posição hierárquica superior (EPPS; BOLAÑOS, 2017).⁵ É nesse quadro que deve ser entendida a vitalidade das línguas e culturas da região. Enquanto uma boa parte da literatura sobre o contato de línguas em ecologias coloniais tende a se concentrar nos efeitos que as línguas europeias tiveram e têm sobre as línguas nativas, é importante lembrar também do papel das relações que se estabelecem no âmbito dos próprios contextos sociais indígenas (EPPS, 2018).

Essa breve introdução sobre esses dois grandes grupos – os Povos do Rio e os Povos da Floresta – serve para ilustrar a dialética entre as compatibilidades e oposições que definem a integração do sistema sociocultural e linguístico do Alto Rio Negro. Isso se verifica mesmo entre as etnias e clãs que podem ser considerados integrantes de um ou outro desses dois grandes grupos.

O Alto Rio Negro se caracteriza como uma ecologia em que há uma ligação estreita entre etnia e língua, e essa associação tem sido crucial para a manutenção do sistema de casamento que existe

5 O contato entre os Povos do Rio com a cultura não indígena se iniciou bem antes do que aquele com os Povos da Floresta, cujas interações mais recentes com o mundo não indígena foram consideravelmente mediadas pelos Povos do Rio. Uma consequência dessa mediação é a de que a imagem que viajantes, exploradores, etnógrafos formaram historicamente dos Povos da Floresta foi moldada pelas perspectivas e interesses dos Povos do Rio (EPPS; BOLAÑOS, 2017).

na região.⁶ Povos da família tukano oriental são exogâmicos, ou seja, eles se casam com pessoas que falam línguas diferentes das suas. A filiação étnica e linguística é feita pela linha paterna: por exemplo, são tukano e falantes de tukano aqueles que têm um pai tukano (EPPS, 2018). Mas, como a prática de casamento é exogâmica, indivíduos tukano normalmente falam também a língua de suas mães. Portanto, é alto o nível de multilinguismo no âmbito da etnia.⁷

Diferentemente, outros grupos que integram os Povos do Rio, como os arawak, seguem um sistema matrimonial distinto, mais próximo ao sistema seguido pelos Povos da Floresta. Eles são endogâmicos, no sentido de que se casam preferencialmente com pessoas que falam a mesma língua; mas mantêm um pouco da prática exogâmica na medida em que se casam com falantes da mesma língua que pertencem a outros grupos e clãs (EPPS; STENZEL, 2013). Ao mesmo tempo, eles se alinham com os tukano (também Povos do Rio) na ornamentação do corpo que caracteriza os rituais masculinos, e no repertório de estilos de dança e de canções que animam esses rituais (HUGH-JONES, 2013).

Uma assimetria semelhante pode ser observada também em meio aos desana, um povo da família tukano que integra o grupo de Povos do Rio. Eles já praticamente preterem o uso de sua língua em favor da língua tukano tal o nível de interação entre as duas etnias (CABALZAR, 2013), mas compartilham com os Povos da Floresta a prática da caça mais intensamente do que outros membros do grupo de Povos do Rio (EPPS; STENZEL, 2013).

A prática exogâmica de casamento exercida por um grande número de grupos étnicos do Alto Rio Negro gera um impacto linguístico ambivalente: ao mesmo tempo em que essa prática pode manter o já mencionado multilinguismo que caracteriza a região, ela pode levar uma língua a entrar em desuso até se extinguir. Chernela (2013) reporta que, na vila de Yapima, os homens e as crianças se comunicam exclusivamente em kotiria, enquanto as mulheres falam desana, kubeo, tukano, além de kotiria, assegurando a continuidade de uma ecologia em que o multilinguismo é ainda bastante vigoroso. Por outro lado, os constantes casamentos entre os tariana e falantes das línguas tukano oriental têm sido considerados a principal causa do estado de quase-extinção daquela língua, na medida em que mulheres tariana tendem a abandonar seu patriléto – a língua falada pela etnia do pai

6 Epps e Stenzel (2013) observam que essa associação entre língua e etnia tem se diluído mais recentemente em função da extinção de algumas línguas e do risco de extinção de outras.

7 Pessoas que falam línguas da família tukano oriental pertencem a aproximadamente quinze diferentes grupos linguísticos. O grupo que fala a língua tukano (uma das línguas da família) é o que conta com o maior número de falantes – mais de seis mil (Chernela 2013). Para mais detalhes sobre a configuração das identidades etnolinguísticas e do multilinguismo na região, ver Stenzel e Khoo (2016) e Silva (2020).

– e adotar a língua da comunidade em que passam a viver depois de casadas (EPPS, 2018).

À primeira vista, não são óbvias as razões que explicam, de um lado, a vitalidade de muitas línguas num determinado ambiente, e, de outro, a troca de uma língua por outra em uma outra ecologia. Fatores de natureza socioeconômica e política devem estar em jogo. Mais adiante volto a falar desses fatores. Mas desde já parece claro que as questões linguísticas que envolvem o Alto Rio Negro configuram-se como um índice inequívoco da complexidade e dinamicidade da língua enquanto um fenômeno sociointeracional, como será descrito a seguir.

3 A língua como um sistema complexo: Breve panorama

Desde o final do século XX, uma vertente de estudos linguísticos tem enfatizado algumas características das línguas humanas que permitem defini-las como sistemas complexos, dinâmicos e adaptativos (BECKNER *et al.*, 2009; ELLIS; LARSEN-FREEMAN, 2009; KRETZSCHMAR, 2015; MUFWENE, 2013; MUFWENE; COUPÉ; PELLEGRINO, 2017; entre muitos outros). Dentre elas, destacam-se:

- seu aspecto interacional: sistemas complexos vivem em permanente troca com outros sistemas da ecologia em que se inserem;
- seu aspecto dinâmico: como consequência de sua contínua interação, as atividades dos sistemas complexos estão sempre sendo coordenadas para se ajustar às contingências do contexto espacial e temporal em que elas se desenvolvem; em consequência, eles estão em permanente mudança;
- seu aspecto não-linear: seu desenvolvimento está atrelado a fatores diversos, que impactam o sistema em diferentes níveis de sua evolução; seus efeitos não são constantes, variando de acordo com as configurações locais e momentâneas em que o sistema se encontra; e
- sua auto-organização: para se manter à beira do caos – uma posição distante de um ponto de quase estagnação de energia, e de outro ponto em que há desordem total – sistemas complexos precisam se reorganizar continuamente.⁸

Tomar a língua como um sistema complexo que exhibe essas características implica assumir que ela é um sistema eminentemente situado; ou seja, não se trata de um sistema autônomo. Como tal, a língua é eminentemente permeável a pressões da ecologia em que vive. Ao considerar a língua como um sistema complexo, a tarefa da linguística passa a abarcar, para além da investigação das interações

⁸ A expressão à beira do caos (*at the edge of chaos*) foi cunhada por Stuart Kauffman (KAUFMANN, 1993, p.174).

internas ao sistema que definem o *valor* de cada uma de suas partes,⁹ a descrição e análise de suas interações externas e dos permanentes processos de auto-organização por que ela passa em resposta às pressões de seu entorno, por meio dos quais novas estruturas emergem.

Tudo isso faz com que as possíveis configurações do sistema linguístico, em sua qualidade de sistema complexo, sejam infinitas e, mais que isso, potencialmente imprevisíveis: sua sensibilidade às condições iniciais de funcionamento (não-linearidade), sua não-autonomia em relação à sua ecologia, seu caráter interacional e a constante auto-organização que garante sua existência resultam em uma indeterminação de sua conformação a cada momento de sua existência. Sistemas complexos são não composicionais, na medida em que as propriedades do todo não são resultantes da soma de suas partes: o comportamento do todo não pode ser inferido a partir do comportamento dos elementos que constituem o sistema. Sendo assim, a teoria da complexidade é também, em grande medida, não reducionista: não é possível entender o comportamento de sistemas complexos pelo comportamento de suas partes, e o comportamento destas por meio do de suas partes e assim sucessivamente. O todo pode apresentar propriedades que não podem ser explicadas pelo comportamento das partes; trata-se de propriedades que emergem do funcionamento coletivo do todo, em um dado momento de sua existência (KAUFMANN, 1995). As interações entre os agentes e as interações do sistema com seu entorno dão origem a propriedades emergentes do sistema. Assim, por pressões externas, itens lexicais e unidades estruturais podem se tornar mais arraigados, ou podem se tornar marginais, vindo a cair em desuso; novas palavras e estruturas se criam, enquanto velhas construções podem vir a ganhar novos significados. De modo mais impactante, dependendo da ecologia de prática linguística, é possível que novas variedades linguísticas emergjam, enquanto outras entram em perigo de extinção.¹⁰

Como dentro da teoria da complexidade a língua é entendida como um fenômeno interacional, não é possível investigar as características da complexidade da língua sem levar em conta a complexidade dos sistemas sociais. É desnecessário dizer que a complexidade que caracteriza as sociedades humanas envolve fatores de ordem diferente daquela que é pertinente para a descrição de fenômenos naturais, sejam físicos ou biológicos. Dentre esses fatores, destacam-se os de natureza

9 A noção de valor é uma das grandes contribuições dos ensinamentos de Saussure em seu *Curso de Linguística Geral*. Cada elemento do sistema linguístico se define pela diferença que existe entre ele e outros elementos do sistema (SAUSSURE, 1969).

10 O modelo teórico conhecido como *Gramática de Construções*, proposto, entre outros, por Goldberg (1995), vem, há tempos, chamando a atenção para o fato de que construções sintáticas têm um significado que transcende a composicionalidade de suas partes. Esse modelo integra a *Linguística Cognitiva*, uma ampla corrente teórica que tem procurado analisar a língua humana a partir de suas características de sistema complexo. Ver, entre outros, Langacker (2008) e Bybee (2010). Para um entendimento de contato e mudança linguística a partir da visão de língua como um sistema complexo, ver Mufwene (2008; 2013; 2018) e Viotti (2013).

ética, econômica, política e cultural, todos necessariamente concebidos em sua dimensão histórica.

Dentro desse quadro, as questões colocadas aqui são:

- i. estudos linguísticos feitos no âmbito da teoria da complexidade devem, em princípio, manter as características de não composicionalidade e não reducionismo que caracterizam os sistemas complexos; será que o uso de modelos computacionais que têm o objetivo de tratar as dinâmicas das ecologias de prática comunicativa têm sido construídos levando em conta a natureza não composicional e não reducionista que caracteriza as interações linguísticas?
- ii. em que medida é pertinente incluir, como fazem Loureiro-Porto e San Miguel (2017, p.187), na própria definição de teoria da complexidade, a restrição de que ela estuda sistemas complexos “*por meio da construção de modelos computacionais*”? Será que os sistemas complexos que envolvem os seres humanos em suas relações sociais e interações linguísticas podem ser bem capturados pela modelagem computacional, e só por ela?
- iii. será que é possível prescindir de relatos etnográficos e descrições sociolinguísticas no trabalho de investigação da complexidade e auto-organização de ecologias de práticas multilíngues?

A seguir, apresento uma síntese de algumas conclusões a que chegaram Loureiro-Porto e San Miguel (2017) sobre as tendências dos falantes em suas escolhas linguísticas em contextos multilíngues, em estudo baseado em modelagem computacional inspirada pela teoria da complexidade, e não em fatos diretamente observados. No item subsequente, discuto as conclusões de Loureiro-Porto e San Miguel, contrastando-as com alguns fatos relatados em estudos etnográficos que descrevem aspectos do multilinguismo que caracteriza a região do Alto Rio Negro.

4 Uma modelagem computacional feita a partir da teoria da complexidade para o estudo da escolha linguística em sociedades multilíngues

Loureiro-Porto e San Miguel (2017) apresentam um modelo de escolha linguística para uma sociedade multilíngue. Alguns fatores que eventualmente possam ter algum papel na sobrevivência de uma ou de várias línguas são levados em consideração. Dentre eles estão (i) a interação de agentes, tomados como entidades autônomas e tomadoras de decisões; (ii) o papel de falantes bilíngues; e (iii) os efeitos que o tipo de rede social em que a escolha é feita pode ter nesse processo.

Os agentes são caracterizados, para os fins do modelo computacional, como homogêneos, podendo ser falantes de uma língua A, de uma língua B, ou das línguas A e B (LOUREIRO-PORTO; SAN MIGUEL, 2017). No que diz respeito aos padrões de interação entre os agentes, para o modelo

usado no estudo são de particular relevância dois princípios: o de homofilia, que determina que são mais prováveis as interações entre agentes que compartilham traços culturais; e o da influência social, que estabelece que interações sociais favorecem similaridade cultural. A interação desses princípios indica as possibilidades de globalização (uniformização) ou de polarização (diversificação) cultural (LOUREIRO-PORTO; SAN MIGUEL, 2017).

A interação dos agentes leva à configuração das redes sociais. O estudo leva em consideração quatro tipos de rede social – as de mundo pequeno (*small world networks*); as redes aleatórias (*random networks*); as redes totalmente conectadas (*fully connected networks*); e as redes com estrutura de comunidade (*networks with community structure*). Como, de acordo com os autores, os três primeiros tipos não são facilmente encontráveis na realidade, para fins da discussão deste artigo vou considerar apenas as redes com estrutura comunitária, que se caracterizam pelo fato de que alguns indivíduos ou grupos de indivíduos da rede são mais conectados entre si do que com outros indivíduos, constituindo, assim, uma estrutura modular (LOUREIRO-PORTO; SAN MIGUEL, 2017).

A simulação computacional feita por Loureiro-Porto e San Miguel compartilha do entendimento explicitado em alguns trabalhos da área segundo o qual modelos computacionais aplicados a questões linguísticas não devem ter o objetivo de formular previsões, mas, sim, o de explorar os possíveis efeitos dos relacionamentos que ocorrem em uma determinada situação real. De forma muito sumária, os resultados obtidos pelo estudo que são de interesse para a discussão teórico-metodológica que quero fazer aqui podem ser agrupados da seguinte maneira: (i) efeitos do bilinguismo; (ii) efeitos do prestígio e da volatividade dos falantes; e (iii) efeitos da estrutura da rede social. Concentro-me nos dois primeiros grupos de resultados, uma vez que, como já dito acima, a única estrutura de rede social que nos interessa é aquela que tem estrutura comunitária. Sendo assim, o último grupo de efeitos não será diretamente levado em consideração.

O estudo de Loureiro-Porto e San Miguel conclui que falantes bilíngues têm um papel central na competição de línguas: em redes com estrutura comunitária o bilinguismo se mostra como um fator que previne ou, pelo menos, retarda a extinção de línguas. Ao mesmo tempo, o estudo mostra que a volatividade dos falantes é um fator mais impactante para a vitalidade das línguas do que o prestígio que algumas podem ter numa determinada ecologia.

Os autores do estudo em discussão não definem o que entendem por ‘prestígio’. O termo, em si, é polissêmico e teria merecido um maior esclarecimento por parte de Loureiro-Porto e San Miguel. É bem verdade que, de modo geral, em discussões superficiais sobre a extinção de línguas, sugere-se

que o ‘baixo prestígio’ de uma língua pode ser considerado uma possível explicação para a decisão dos falantes de abandonar seu uso. O estudo de Loureiro-Porto e San Miguel mostra, diferentemente, que o peso dado ao ‘prestígio’ como um fator crucial para a manutenção de línguas – mesmo que sua definição não tenha sido dada – é, por vezes, excessivo. Segundo os autores, o que é verdadeiramente determinante é a volatilidade dos falantes, ou seja, sua disposição para mudar de língua: quanto menos os falantes estão dispostos a usar outra língua, tanto maior a chance de suas línguas sobreviverem. Mesmo assim, o estudo mostra que, em redes como as que têm estrutura de comunidade, certos tipos de interações locais podem levar ao aumento do papel do ‘prestígio’ como fator responsável pelas escolhas linguísticas dos falantes. Seja como for, além de não esclarecer o que se deve entender como ‘prestígio’, o trabalho de Loureiro-Porto e San Miguel tampouco sugere qualquer explicação para a possível disposição dos falantes de abandonar suas línguas nativas em favor de outras. Mufwene (2003) adverte que conceitos como o de ‘prestígio’ requerem uma discussão mais cuidadosa e que a volatilidade dos falantes não pode ser tomada como um fator geral, na medida em que ela é sempre motivada. Mais adiante, volto a discutir esses conceitos para apontar como eles são sensíveis às características específicas de cada ecologia multilíngue.

Loureiro-Porto e San Miguel concluem realçando o ponto de vista estritamente teórico de seu estudo. O fato de ele revelar o impacto que bilinguismo, prestígio de línguas e volatilidade dos falantes (associados a configurações das redes de prática social) podem ter na vitalidade linguística é, segundo os autores, uma evidência de que estudos teóricos podem expor alguns aspectos do processo de competição entre línguas que análises de dados não apontariam. Mesmo assim, os autores afirmam que seu estudo precisa ainda ser validado, e isso só pode ser feito por meio de uma comparação com dados etnográficos. No item seguinte, discuto o modelo de Loureiro-Porto e San Miguel à luz de trabalhos sociolinguísticos e de algumas etnografias feitas no Alto Rio Negro.

5 Modelos computacionais versus etnografias

Como seria esperado, em várias passagens de seu texto, Loureiro-Porto e San Miguel enfatizam a importância da modelagem computacional para o entendimento da evolução das línguas e de sua vitalidade. Para eles, mesmo uma simulação bastante simplificada se configura como um avanço em relação a discussões sobre as condições necessárias para explicar o processo de evolução linguística, e podem contribuir para questionar as crenças de senso comum dos pesquisadores cujas análises se baseiam na observação dos fatos (LOUREIRO-PORTO; SAN MIGUEL, 2017, p. 191, 196; 212).¹¹

¹¹ Especificamente, Loureiro-Porto e San Miguel defendem a simplicidade dos modelos matemáticos pelos seguintes motivos: (i) “eles evitam a miríade de variáveis com os quais estudos sociológicos e humanísticos tipicamente se veem às voltas” (p. 191); (ii) eles propiciam a colaboração entre especialistas em modelagem computacional e pesquisadores

Parece-me indiscutível a contribuição que a modelagem computacional pode dar para algumas áreas da linguística, como aquelas que trabalham com grandes bancos de dados, com o processamento e a aquisição da linguagem, entre outras. Entretanto, como os próprios Loureiro-Porto e San Miguel apontam, modelos matemáticos usados no âmbito da teoria da complexidade precisam ser simples: eles devem abranger um número limitado de componentes, e evitar lidar com as incontáveis variáveis com as quais os estudos das humanidades se veem às voltas (LOUREIRO-PORTO; SAN MIGUEL, 2017). Se é assim, a questão é: será que estudos que dizem tomar a teoria da complexidade como base, mas que, ao mesmo tempo, acatam a simplificação imposta pelo uso de modelos computacionais – aqui exemplificados pelo artigo de Loureiro-Porto e San Miguel – podem, de fato, trazer alguma contribuição para o entendimento de ecologias multilíngues, que são verdadeiros sistemas complexos, auto-organizados e adaptativos? Será que o uso de ferramentas computacionais, talvez pertinente para a modelagem de sistemas complexos que não podem ser diretamente observados (como é o caso de muitos fenômenos físicos, químicos e biológicos), é totalmente adequado para a descrição de sistemas complexos de natureza social, cultural e linguística?

O texto de Loureiro-Porto e San Miguel leva ao entendimento de que a aplicação de modelos computacionais a fenômenos sociais e linguísticos geram uma tensão: por um lado, a língua humana entendida como um sistema complexo é fundamentalmente tomada como interacional, não composicional, e caracterizada pela sempre constante e imprevisível emergência de padrões globais resultantes da manipulação do sistema pelos falantes como respostas às pressões externas da ecologia em que ela vive; por outro, com o intuito de que essas características da complexidade sejam explicitadas de uma maneira supostamente mais objetiva do que aquela feita por descrições sociolinguísticas e etnográficas, a língua, em seus aspectos sociolinguísticos, precisa ser concebida de maneira simplificada, limitada em sua variabilidade, reduzida a um modelo que parece não poder capturá-la em toda a sua inteireza e dinamicidade. Isso sugere que, para que modelos computacionais possam trazer resultados de interesse para a pesquisa linguística, grande parte daquilo que caracteriza a língua como um sistema complexo precisa ser deixada de lado.

Além da simplificação exigida pela modelagem computacional que elimina ou neutraliza muito daquilo que faz da língua um sistema complexo, modelos computacionais parecem ser antes determinados apriorística e genericamente, em vez de ser determinados pelas características concretas

das áreas humanas que “geralmente não são treinados para a análise matemática de dados quantitativos” (p. 196); e, (iii) “modelos teóricos não baseados em dados podem contribuir para *questionar as crenças de senso-comum de um observador*” (p.212). A tradução e os grifos são meus. Com certeza, existe um fundo de verdade no que dizem Loureiro-Porto e San Miguel nesses trechos; entretanto, as afirmações dos autores, tomadas em sentido geral, não parecem fazer justiça a um grande número de pesquisas de base descritiva feitas no âmbito das ciências sociais e da linguística.

de um sistema sociolinguístico particular. O estudo de Loureiro-Porto e San Miguel, por exemplo, avalia, dentre outros fatores que podem ter impacto na vitalidade das línguas em ecologias multilíngues, o ‘prestígio’ de algumas línguas e a disposição dos falantes de abandonar suas línguas nativas em favor da língua de prestígio (volatilidade). Entretanto, primeiramente, como já visto, esses fatores são excessivamente gerais e carecem de uma definição mais específica para poderem ser efetivamente considerados definidores da configuração de uma determinada ecologia linguística. Além disso, eles são tomados como fatores macrossociais passíveis de condicionar a dinâmica linguística de contextos sociais considerados genericamente, ou seja, sem que aspectos microssociais, que mais precisamente definem uma ecologia de prática linguística, tenham sido levados em conta.

A consideração das características microssociais de cada sociedade multilíngue é fundamental para o entendimento das possibilidades de manutenção ou extinção de línguas. A partir da observação e análise da dinâmica sócio-histórica e econômica de várias ecologias multilíngues, Mufwene (2003, 2004, 2008) registra que o papel de fatores como a educação, a imprensa, e o poder e o prestígio social de alguns falantes podem não ser tão significativos para a vitalidade linguística como se imagina. O que é verdadeiramente determinante para as escolhas dos falantes é o quanto adotar uma língua diferente de sua língua nativa pode se mostrar vantajoso para eles em termos econômicos e sociais; e isso varia não só historicamente em uma mesma ecologia, mas também, e muito, de ecologia para ecologia (MUFWENE, 2003; 2004; 2008). Surpreende, portanto, que um fator tão crucial para o entendimento de sistemas ecológicos multilíngues e da vitalidade linguística como esse não tenha sido escolhido para ser avaliado pela modelagem feita por Loureiro-Porto e San Miguel.

A análise de Mufwene sobre a vitalidade das línguas em ecologias multilíngues mostra, ainda, que a volatilidade dos falantes não é imotivada. Ela está sempre ligada aos benefícios que um indivíduo pode ter ao falar a língua associada às redes de poder econômico, político e social. No Alto Rio Negro, esse fator é central para a explicação de muitos cenários. Especialmente entre a região do povoado de Iauaretê (na confluência dos rios Vaupés e Papuri, fronteira com a Colômbia) e a zona urbana do município de São Gabriel da Cachoeira, o poder dos povos falantes da língua tukano é bastante grande; sua língua tem sido considerada a língua franca da região desde o início do século XX, suplantando o status de língua franca do nheengatu, língua da família tupi-guarani levada para a região pelos colonizadores durante o século XIX (EPPS; STENZEL, 2013; EPPS; BOLAÑOS, 2017; EPPS, 2018). Tukano é a língua mais falada hoje na região do Vaupés brasileiro, já sendo considerada como a primeira língua de grande parte dos desana e miriti-tapuya, e tem sido adotada por outros povos como os tariana, arapaso e wa'ikhana, cujas línguas ou já se perderam, ou estão em perigo de se extinguir. Isso revela, de um lado, o poder político dos tukano (CABALZAR, 2013), alavancado não

só pelo seu predomínio populacional, como também pelo incentivo dos missionários salesianos que chegaram à região há mais de um século (CHERNELA, 2013); e, de outro, o fato de que, ao escolher tukano como sua nova língua, indivíduos que vêm de grupos falantes de outras línguas se aproximam dos detentores do poder, vindo, possivelmente, a obter vantagens econômicas e sociais. Uma senhora baré, habitante da zona urbana de São Gabriel da Cachoeira, disse que seus filhos não querem mais falar nheengatu, sua língua nativa: eles preferem falar português, a língua que vai ajudá-los a conseguir melhores empregos na cidade; e querem falar tukano, que é a língua que, ao ver deles, ratifica sua identidade indígena, garantindo-lhes o acesso aos benefícios da FUNAI (Fundação Nacional do Índio) (VIOTTI, 2017). Pode-se dizer que se reitera hoje um antigo desconforto registrado desde a instauração das primeiras escolas missionárias salesianas, que privilegiavam o uso da língua tukano em prejuízo de outras línguas da região. Como relata Cabalzar, alunos das escolas missionárias que falavam outras línguas que não o tukano sentiam vergonha de usar suas línguas nativas por causa da pressão exercida pelos padres (CABALZAR, 2013). Diante desse quadro, como é que poderiam ser entendidas as noções de ‘prestígio’ e de volatividade dos falantes? Para alguns fins, o ‘prestígio’ é atribuído ao português; para outros, à língua tukano. No que diz respeito à disposição dos falantes em abandonar a língua de sua etnia para adotar outras não é gratuita: ela é determinada pelos ganhos socioeconômicos e políticos que essa troca de línguas pode gerar. Essas questões são específicas dessa ecologia e não podem ser facilmente generalizadas e transplantadas para explicar o que acontece em outros contextos sociolinguísticos, como vai ser visto mais adiante.

Um dos efeitos do poder dos tukano na política linguística da região do Alto Rio Negro foi a oficialização de sua língua no município de São Gabriel da Cachoeira. É verdade que tukano é uma dentre quatro línguas oficiais do município, acompanhada do português, do nheengatu e do baniwa. Entretanto, quando a questão da educação indígena entra em cena, o papel da língua tukano ganha ainda mais espaço. A legislação brasileira garante aos povos indígenas o direito a uma educação feita em suas línguas nativas. Existem políticas de treinamento de professores indígenas, de incentivo à publicação de material didático especialmente desenvolvido para a educação indígena, e de estímulo a encontros para a elaboração de currículos especiais que respeitem as tradições e o tipo de vida indígena. Isso tudo é um grande avanço, mas a realidade é bem diferente: esses são projetos de longo prazo que requerem verbas vultuosas; e, infelizmente, como se sabe, a educação não tem sido uma das prioridades das políticas governamentais brasileiras.

A questão que nos interessa aqui é que o Alto Rio Negro abrange um território extremamente vasto, em que mais de vinte línguas são faladas. No núcleo urbano de São Gabriel da Cachoeira, onde estudam crianças e jovens das mais variadas etnias, a única língua usada nas escolas é o português.

Nheengatu é ensinado, mas em uma carga horária de apenas uma hora por semana (SCHULIST, 2018). Nas áreas rurais, é difícil que haja professores de todas as etnias, falantes de todas as línguas para que a educação possa ser feita na língua de cada criança e jovem. Como a presença dos tukano na região é massiva, muitas das escolas indígenas têm professores que dão aulas em língua tukano, e há falta de professores nativos de outras línguas. Um caso ilustra bem a situação: uma jovem de 20 anos da etnia hup, que morava e frequentava a escola indígena na aldeia de Taracúá Igarapé, reporta que suas aulas eram quase sempre dadas em língua tukano, algumas vezes em português. Isso acontecia porque o número de professores da etnia hup é muito pequeno para dar conta da educação de todas as crianças e jovens falantes de hup. Como a aluna da escola de Taracúá Igarapé não fala tukano e fala pouco português, sua educação ficou bastante prejudicada. A alternativa que ela teria para obter mais conhecimentos seria aprender a falar tukano, uma opção que muitos acabam seguindo (VIOTTI, 2017).

A questão de ‘prestígio’ linguístico e de volatividade dos falantes deve, portanto, ser sempre posta em perspectiva. Mufwene mostra, por exemplo, que, ao longo da história dos contatos entre povos e línguas, o ‘prestígio’ atribuído a línguas que contavam com um sistema de escrita em oposição a línguas ágrafas, e o ‘prestígio’ atribuído a uma determinada língua porque ela é a língua usada no processo de escolarização quase nada tiverem ou têm a ver com a vitalidade das línguas, com sua extinção, ou com a opção dos falantes por outras línguas. Entretanto, o autor observa que esses fatores podem eventualmente estar associados a um outro fator, qual seja, o de que essa língua oferece, a seus falantes, melhores meios para o desenvolvimento de sua vida num contexto socioeconômico específico. Nesses casos sim, a escrita e a associação de uma língua ao sistema educacional podem ser considerados como relevantes para o entendimento da dinâmica da competição entre línguas em uma determinada ecologia, fundamentando as escolhas dos falantes por uma língua, e não outra (MUFWENE, 2003).

Além disso, Mufwene aponta uma outra questão relacionada ao ‘prestígio’ que mostra que esse conceito necessita ser entendido de maneira precisa. Há variedades linguísticas, como o inglês afro-americano, nos Estados Unidos, e vários crioulos ao redor do mundo que não parecem estar em risco de extinção, e convivem bastante bem com as variedades consideradas de ‘maior prestígio’, quer porque sejam variedades ligadas ao sistema educacional, quer porque sejam atreladas a maiores vantagens socioeconômicas e políticas. O que explica essa convivência é que há uma complementaridade entre as funções das duas variedades: enquanto algumas funções socioeconômicas exigem o uso da variedade considerada de ‘prestígio’, a socialização dos falantes com seus pares é feita preferencialmente na variedade não prestigiada. Isso sugere que a coexistência social e geográfica

de variedades de maior ou menor ‘prestígio’ não leva necessariamente à necessidade de uma escolha entre elas se elas se prestam a diferentes fins (MUFWENE, 2003).

Essa observação nos leva a pensar que também o bilinguismo, que de acordo com o estudo de Loureiro-Porto e San Miguel deve ser considerado um possível fator de favorecimento à manutenção de línguas ou à sua perda, precisa ser bem caracterizado.¹² O que foi descrito acima sobre a convivência de variedades que, por qualquer motivo, sejam prestigiadas, com variedades estigmatizadas acontece também com línguas diferentes: duas ou mais línguas se mantêm em boa convivência se cada uma delas servir a uma função diferente.¹³ Dependendo da ecologia, outros fatores não associados a prestígio social e político podem também entrar em jogo. No continente africano, por exemplo, nos territórios que foram colonizados segundo o modelo de exploração,¹⁴ a organização socioeconômica é tal que várias línguas convivem, cada uma associada a um subsistema social: um subsistema minoritário de trabalhadores de empresas, escritórios, órgãos públicos, em que se usa uma língua europeia; e um subsistema majoritário de trabalhadores braçais responsáveis pelos serviços agrícolas, de mineração, de construção, etc., em que se usa alguma língua franca, como por exemplo o swahili, falado na África oriental. Mas fora do ambiente profissional, mesmo a elite africana, que aprendeu e domina a língua europeia, conhece e usa a língua franca local para se comunicar com a maioria da população, e preserva suas línguas étnicas, usando-as nos encontros sociais de família e de amigos (MUFWENE, 2003).

Sendo assim, em um tipo de ecologia como aquela que caracteriza grande parte da África subsaariana, o bilinguismo (ou o multilinguismo) de fato parece não colocar em grande risco as

12 Como já visto, Loureiro-Porto e San Miguel (2017) concluem, com base em seu modelo computacional, que a presença de bilíngues em redes com estrutura de comunidade previne a extinção de línguas; em redes totalmente conectadas – aquelas em que todos interagem com todos – a presença de bilíngues reduz a possibilidade de coexistência de línguas; e em redes de mundos pequenos – aquelas em que novos conhecidos têm amigos em comum – agentes bilíngues aceleram a morte de línguas (LOUREIRO-PORTO; SAN MIGUEL, 2017).

13 Se pensarmos no caso do filho da senhora baré descrito acima, sua intenção é, em princípio, a de usar português para uma função e tukano para outra. Enquanto essa complementariedade de funções puder ser mantida, é possível que ele se configure como um indivíduo bilíngue.

14 Em uma série de publicações, Mufwene vem enfatizando que as ecologias de uso linguístico que emergiram do processo de colonização europeia iniciado no século XV têm suas configurações atreladas ao padrão de colonização utilizado em cada território. Inicialmente, o padrão mais utilizado foi o de povoamento: os europeus se estabeleciam nas colônias com o intuito de fazer delas seu novo lar. Esse foi o tipo de colonização característico das Américas. De um ponto de vista linguístico, as línguas europeias se tornaram os vernáculos dessas colônias. Em um segundo momento da expansão imperial europeia, a colonização passou a utilizar um padrão de exploração. Esse foi o padrão caracteristicamente usado na colonização da porção subsaariana do continente africano. Por esse modelo, os colonizadores criaram uma estrutura sociopolítica e econômica seguindo um modelo europeu paralela à estrutura africana (MUFWENE, 2001; 2003; 2004; 2008).

línguas étnicas ancestrais.¹⁵ Entretanto, no caso das línguas nativas americanas como as brasileiras por exemplo, a situação é bastante diferente. O sistema de colonização utilizado pelos portugueses no Brasil foi o de povoamento.¹⁶ Historicamente uma das consequências desse tipo de colonização tem sido a troca de línguas, com a consequente extinção de algumas, e a apropriação das línguas da população econômica e politicamente dominante pelos grupos de menor poder. Essa apropriação é um dos grandes fatores responsáveis pela emergência de novas variedades de uma língua europeia nas antigas colônias das Américas e pela extinção de várias línguas nativas ancestrais (MUFWENE, 2003; 2004; 2008).

No território brasileiro, a estimativa é a de que mais de mil línguas nativas fossem faladas à época da chegada dos europeus; no início do século XX, esse número girava em torno de 180 (RODRIGUES, 1993; LEITE; FRANCHETTO, 2006). No Alto Rio Negro, apesar de felizmente haver ainda várias línguas que contam com um razoável número de falantes, a troca de línguas acontece a olhos vistos: os baré, já há algum tempo, deixaram sua língua da família arawak para falar nheengatu; os tariana, que participam integralmente do sistema cultural tukano, também praticamente já quase não usam mais sua língua arawak preferindo a língua tukano; os desana, kotiria e wa'ikhana, todos originalmente grupos falantes de línguas da família tukano oriental, estão, em muitas situações, deixando de falar suas línguas étnicas para dar preferência à língua tukano. Naturalmente, todos aqueles que têm algum tipo de contato com os centros urbanos também falam português.

Como se vê então, parece correta a observação de Mufwene de que, em sociedades multilíngues, o que está por trás do favorecimento de algumas línguas em detrimento de outras são as pressões ecológicas: a absorção de falantes de uma língua pelo sistema cultural, social, econômico ou político de falantes de outra faz com que, aos poucos, os falantes que se integraram a esses outros sistemas percam o interesse em manter suas línguas originais. No Alto Rio Negro, em especial na bacia do rio Vaupés e no centro urbano de São Gabriel da Cachoeira, o sistema indígena que parece mais atraente para os habitantes da região é o da etnia tukano, que, como já visto, não só é demograficamente dominante, mas também detém grande poder político e econômico. Um círculo vicioso se cria: por essas razões, tukano é a língua franca dessa região e é uma das línguas oficiais do município de São Gabriel da Cachoeira; ao ser uma língua franca e uma língua oficial, a língua tukano acaba por

15 Entretanto, as línguas francas – aquelas usadas nas atividades econômicas que atraem mais pessoas – têm registrado, como esperado, um aumento no número de falantes. Elas têm, inclusive, se tornado a língua nativa da maioria das crianças nascidas nos centros urbanos (MUFWENE, 2003). Portanto, se algumas línguas étnicas estão em perigo de extinção, isso se deve ao maior interesse que as pessoas têm tido em falar as línguas francas, justamente aquelas que estão associadas a vantagens socioeconômicas acessíveis a uma grande parte da população.

16 Ver nota 14.

ganhar uma força ainda maior. Indivíduos bilíngues, falantes de suas línguas ancestrais e de tukano, podem, com o tempo, abandonar as primeiras e manter a segunda, mais vantajosa para eles em todos os sentidos. É possível ainda que, em algum momento no futuro, até mesmo a língua tukano deixe de ser interessante para os habitantes da região, que talvez venham a preferir falar exclusivamente o português. Essa tendência já se verifica nos centros urbanos. Schulist (2018) reporta que famílias kotiria vivendo em São Gabriel da Cachoeira, apesar de falantes da língua, tendem a falar português com os filhos e entre si. O mesmo acontece com um grande número de barés que vivem na cidade, que já usam pouco o nheengatu, preferindo usar o português em suas interações comunicativas com outros barés, e que não têm passado a língua para seus filhos e netos (VIOTTI, 2017). Diferentemente, e como evidência da complexidade que caracteriza contextos multilíngues, um pai da etnia hup expressa uma opinião sobre a educação de seus filhos, que enfatiza a necessidade de complementaridade entre as funções de cada língua, intuitivamente capturando a ideia de que uma das chaves para a manutenção de várias línguas da ecologia multilíngue é assegurar que cada uma seja usada para diferentes fins. Esse pai se mostra surpreendentemente cético sobre a política de educação indígena. Para ele, a escola é uma instituição de brancos. Ele quer que seus filhos vão à escola para aprender português e as práticas do mundo dos brancos: a importância dos documentos, como consegui-los, suas funções, como resolver problemas no banco, etc. A seu ver, não é necessário que seus filhos tenham aulas *de* língua hup. A língua, os mitos, as práticas culturais hup, eles aprendem em casa, nas aldeias, nos clãs, no seio da vida indígena.¹⁷

Com certeza, fatores sociais concebidos de uma perspectiva bem ampla e geral, como bilinguismo, prestígio da língua e volatividade dos falantes podem ser considerados elementos que afetam as escolhas linguísticas dos indivíduos que participam de ecologias multilíngues. Mas o que os relatos de base etnográfica acima revelam é que há muitos outros fatores em jogo que determinam o comportamento das três condições escolhidas por Loureiro-Porto e San Miguel. E justamente é da ação conjunta desses outros fatores que emergem as características que fazem da língua em interação um sistema complexo, dinâmico e adaptativo.

17 Essa informação me foi passada, em comunicação pessoal, pelo antropólogo Prof. Dr. Danilo Paiva Ramos, pelo que agradeço. O ponto de vista desse pai hup ecoa uma reflexão mais ampla feita pelos povos hupd'äh e yuhupd'äh de que a escola é um projeto de colonização que, como tudo na empresa colonial, tende a desconsiderar os saberes e os modos de vida indígenas. Para muitos, a escola afasta a criança e o jovem do ensinamento tradicional que é passado pelos avós e pelos pais. Os jovens perdem o interesse por aprender o que os velhos poderiam lhes ensinar; os velhos acabam morrendo e, com eles, morre também muito de sua cultura ancestral. Apesar de terem certo apreço pela educação indígena, que tem o objetivo de respeitar as tradições de cada etnia, os povos hupd'äh e yuhupd'äh não deixam de comentar os efeitos que a escolarização tem tido em sua organização sociocultural. A esse respeito, consultar o site da FOIRN – Federação das Organizações Indígenas do Rio Negro: <https://foirn.org.br>; <https://foirn.blog/2014/05/09/povos-hupdah-e-yuhupdeh-discutem-educacao-indigena-e-educacao-escolar-indigena-em-v-seminario-local/>.

Ecologias multilíngues não envolvem apenas questões como a manutenção ou a perda de línguas, mas também a possibilidade de desenvolvimento de novos sistemas sociolinguísticos. Um fator que não consta do estudo de Loureiro-Porto e San Miguel, por exemplo, é o que diz respeito à emergência de diferentes variedades linguísticas como consequência do contato entre as línguas em certos tipos de ecologias bi- ou multilíngues. Recentemente, estudiosos do contato de línguas têm observado a emergência daquilo a que Lüpke se refere como *fala multilíngue* ('multilingual speech') (LÜPKE 2016). A fala multilíngue emerge em ecologias em que a interação não é configurada a partir de domínios de especialização ou funções específicas associadas a cada uma das línguas, nem por relações hierárquicas (ou de 'prestígio' – qualquer que seja a aceção atribuída a esse termo) entre as diferentes línguas. Apesar de falantes e pesquisadores que estudam as práticas linguísticas assumirem que cada falante privilegia, por diversas razões, uma determinada língua e a usa *essencialmente* – ou seja, sem misturá-la com outras línguas que participam da mesma ecologia – é preciso ter sempre em conta que os sistemas linguísticos não são estanques e impenetráveis. Eles são permeáveis, variáveis e dinâmicos (LÜPKE, 2016; STENZEL; KHOO, 2016). Em contextos multilíngues, os falantes, em muitos casos, fazem uso das diferentes possibilidades linguísticas que têm a seu dispor para criar não uma identidade linguística única e estática, mas identidades alternativas, sensíveis ao contexto e administradas localmente, fazendo emergir, nas interações, as variações linguísticas que se adequam melhor às suas necessidades em diferentes momentos. Em alguns casos, em uma mesma prática conversacional, diferentes assuntos são tratados por meio de diferentes línguas como uma estratégia social que permite aos falantes marcar variadas filiações identitárias, conforme seja necessário (LÜPKE, 2016). Em outros casos, essas variações se caracterizam por ser híbridas, no sentido de que envolvem o uso de itens lexicais ou morfemas gramaticais de línguas diferentes em um mesmo enunciado (STENZEL; KHOO, 2016).¹⁸

18 Stenzel e Khoo (2016) apresentam e analisam práticas discursivas dos kotiria, em que, falantes mantêm sua filiação primária ao uso da língua de sua etnia determinada patrilinearmente, ao mesmo tempo em que empregam itens lexicais do português para fazer referência a entidades e elementos não indígenas. Esses termos, no entanto, são perfeitamente integrados à estrutura gramatical da língua kotiria. Ao mesmo tempo, no caso de uma falante cuja mãe é tariana, falante de tukano, as autoras observaram o uso de alternância no uso de formas semanticamente equivalentes dos sistemas pronominais de kotiria e tukano. Essa alternância é significativa no discurso analisado. O uso da forma exclusiva do pronome de 1ª pessoa do plural em kotiria tinha como referente um grupo menor de pessoas a que a falante se referia, enquanto o pronome do tukano tinha como referente um grupo maior de pessoas, que incluía o grupo menor referido pela forma kotiria. Como explicam Stenzel e Khoo, essa alternância não só permite a referência a dois grupos 'exclusivos' de pessoas – um grupo inserido no outro – como também é crucial para a mudança de perspectiva narrativa. Na mesma direção, Silva (2020) reporta casos de hibridismos linguísticos que ocorrem em interações entre falantes de línguas da família tukano oriental que têm diferentes filiações etnolinguísticas. O trabalho investiga interações ocorridas em uma comunidade desano-siriana no território colombiano, em que vivem, além de desanos e sirianos, alguns indivíduos etnolinguisticamente caracterizados como tukano, kotiria, kubo e yuriti. Todas as interações analisadas apresentam intensa alternância de códigos, em que os falantes tendem a usar a língua com que se identificam, e não a língua desano, que é a língua que segue a linha da patrilinearidade, normalmente usada em contextos mais formais.

Apesar de questões como o hibridismo linguístico, que se verifica em práticas conversacionais em contextos multilíngues, não terem sido objeto do estudo de Loureiro-Porto e San Miguel (2017), elas evidenciam, uma vez mais, a complexidade dos fatores que caracterizam o multilinguismo em determinados contextos sociais, enfatizando a importância de relatos detalhados dos contextos sociais e das interações linguísticas, para a descrição e o entendimento de como o contato linguístico se dá e de suas consequências para a troca de línguas, para a mudança e para a vitalidade linguísticas. Como afirma Lüpke, qualquer abordagem que pretenda dar conta de contextos multilíngues e que não leve em conta as intenções dos falantes e sua prática situada tende ao insucesso (LÜPKE, 2016).

Considerações finais

Com base no que foi dito acima, fecho este texto resumindo as respostas às perguntas instigadas pelo trabalho de Loureiro-Porto e San Miguel (2017) (ver item 3 acima). O objetivo dessas questões foi o de problematizar a adequação de modelos computacionais para o tratamento das características das interações humanas entretidas como sistemas complexos, cotejando-os com relatos etnográficos e descrições sociolinguísticas.

A teoria da complexidade parece ser uma alternativa promissora para ampliar os horizontes da investigação sobre a língua humana. Adotá-la, no entanto, parece implicar cuidados no sentido de evitar a ênfase em tudo o que possa ser tomado como tendências gerais que envolvem o funcionamento da linguagem, desvinculando-as de sua dinamicidade e de suas especificidades contextuais. Afinal, o ponto de vista que a teoria da complexidade indica para os estudos da língua humana vem justamente enfatizar a dinamicidade do uso e da prática linguística no processo de construção das interações humanas. Ao tomar os falantes como homogêneos e ao determinar a priori quais os fatores que vão constituir o modelo sem uma observação prévia de uma determinada ecologia (ver item 4 acima), os modelos computacionais acabam por criar uma versão idealizada e descontextualizada de língua, afastando-a da complexidade sociointeracional que a caracteriza. Em contextos multilíngues caracterizados pelo contato de várias línguas, essa complexidade se revela por diversas formas, uma das quais a de que ecologias sociais aparentemente idênticas podem fazer emergir diferentes estruturas linguísticas e sociolinguísticas (TRUDGILL 2011; LÜPKE 2016).

Com relação à sugestão de Loureiro-Porto e San Miguel de que a modelagem computacional é uma ferramenta de grande importância para a investigação de sistemas complexos (LOUREIRO-PORTO; SAN MIGUEL, 2017, p. 187; 195), talvez seja mais pertinente pensar nos modelos e simulações computacionais como uma alternativa interessante para o estudo dos sistemas complexos que

envolvem seres humanos em suas relações sociais. A partir do estudo de Loureiro-Porto e San Miguel aqui cotejado com fatos de base etnográfica e sociolinguística, fica claro que, para que esses modelos efetivamente venham a oferecer uma complementação para estudos feitos no âmbito da etnografia, da linguística antropológica e da sociolinguística, eles não podem abrir mão de computar um conjunto de fatores de natureza diversa que refletem a complexidade que caracteriza contextos multilíngues.

Como já mencionado, é possível que sistemas complexos de natureza física, química ou biológica, que não podem ser integralmente observados na dinamicidade de seus funcionamentos, dependam de modelagem computacional para ser estudados. No caso da língua tomada como sistema complexo em toda a sua inteireza sociointeracional, um sistema que é usado e manipulado por seres humanos, que são, na verdade, os agentes da emergência da complexidade, as modelagens computacionais parecem acabar por ser bastante reducionistas colocando-se, assim, em desacordo com a teoria da complexidade.

Mais de cem anos após o estabelecimento da linguística como ciência autônoma, ela se vê em um momento propício para expandir seus horizontes, encarando a língua como uma forma de vida, que compartilha com outras formas de vida as características de ser um sistema complexo, dinâmico, aberto, auto-organizado, situado sociocultural e historicamente. Abre-se agora uma oportunidade para a linguística passar a investigar a língua a partir da premissa de que ela é um sistema em fluxo constante, com novos elementos e propriedades emergindo a cada momento, enquanto outros se modificam ou desaparecem. Esse me parece ser um grande desafio. Contextos multilíngues oferecem dados ricos e variados para essas novas investigações. Seria interessante que especialistas da computação se juntassem a linguistas descritivistas, linguistas antropólogos e sociolinguistas na empreitada de capturar tão detalhadamente quanto possível as características da língua em ação e em seu habitat natural, que é a vida social. Só assim modelos computacionais podem vir a dar uma contribuição verdadeiramente substancial para o entendimento da dinâmica social, econômica, política e histórica que explica a vitalidade das línguas em ecologias multilíngues e a miríade de formas que elas assumem.

REFERÊNCIAS

ATHIAS, Renato. *Hupdë-Maku et Tukano: relations inégales entre deux sociétés du Uaupés amazonien* (Brésil). 1995. Tese de Doutorado. Université de Paris X.

BECKNER, Clay; BLYTHE, Richard; BYBEE, Joan; CHRISTIANSEN, Morten H.; CROFT,

William; ELLIS, Nick C.; HOLLAND, John; KE, Jinyun; LARSEN-FREEMAN, Diane; SCHOENEMANN, Tom. Language is a complex adaptive system. *Language Learning*, Ann Arbor, vol. 59, n.1, p. 1-26, dezembro 2009.

BYBEE, Joan L. *Language, usage and cognition*. Cambridge/New York: Cambridge University Press, 2010.

CABALZAR, Aloísio. Organização socioespacial e predomínios linguísticos no rio Tiquié. In: EPPS, Patience L.; STENZEL, Kristine (eds.). *Upper Rio Negro*. Cultural and linguistic interaction in Northwestern Amazonia. Rio de Janeiro: Museu do Índio (FUNAI)/Museu Nacional, 2013. p. 129–162.

CHERNELA, Janet M. Toward an East Tukano ethnolinguistics: Metadiscursive practices, identities, and sustained linguistic diversity in the Vaupés basin of Brazil and Colombia. In: EPPS, Patience L.; STENZEL, Kristine (eds.). *Upper Rio Negro*. Cultural and linguistic interaction in Northwestern Amazonia. Rio de Janeiro: Museu do Índio (FUNAI)/Museu Nacional, 2013. p.197–244.

ELLIS, Nick C.; LARSEN-FREEMAN, Diane. *Language as a complex adaptive system*. Oxford: Wiley Blackwell, 2009.

EPPS, Patience L. The Vaupés melting pot: Tucanoan influence on Hup. In: AIKHENVALD, Alexandra Y.; DIXON, Robert M.W. (eds.). *Grammars in contact: a cross-linguistic typology*. Oxford/New York: Oxford University Press, 2007. p. 267-289.

EPPS, Patience L. Contrasting linguistic ecologies: Indigenous and colonially mediated language contact in northwest Amazonia. *Language and Communication*, vol. 62, p. 156-169, 2018.

EPPS, Patience L.; STENZEL, Kristine. Introduction: Cultural and linguistic interaction in the Upper Rio Negro region. In: EPPS, Patience L.; STENZEL, Kristine (eds.). *Upper Rio Negro*. Cultural and linguistic interaction in Northwestern Amazonia. Rio de Janeiro: Museu do Índio (FUNAI)/Museu Nacional, 2013. p.13-50.

EPPS, Patience L.; BOLAÑOS, Katherine. Reconsidering the “Makú” language family of Northwest Amazonia. *International Journal of American Linguistics*, vol. 83, n. 3, p. 467–507, 2017.

GOLDBERG, Adele E. *A construction grammar approach to argument structure*. Chicago: The University of Chicago Press, 1995.

HALE, Ken; KRAUSS, Michael; WATAHOMIGIE, Lucile J.; YAMAMOTO, Akira Y.; CRAIG, Colette; JEANNE, LaVerne M.; ENGLAND, Nora C. Endangered languages. *Language*, Ann

Arbor, vol. 68, n.1, p. 1–42, 1992.

HUGH-JONES, Stephen. Pandora's box - Upper Rio Negro style. In: EPPS, Patience L.; STENZEL, Kristine (eds.). *Upper Rio Negro*. Cultural and linguistic interaction in Northwestern Amazonia. Rio de Janeiro: Museu do Índio (FUNAI)/Museu Nacional, 2013. p.53-90.

JACKSON, Jean E. *The fish people*. Linguistic exogamy and Tukanoan identity in Northwest Amazonia. Cambridge: Cambridge University Press, 2012.

KAUFFMAN, Stuart A. *At home in the universe*. The search for the laws of self-organization and complexity. Oxford: Oxford University Press, 1995.

KRETZSCHMAR JR., William A. *Linguistics and complex systems*. Cambridge: Cambridge University Press, 2015.

LANGACKER, Ronald W. *Cognitive Grammar: A Basic Introduction*. Oxford/New York: Oxford University Press, 2008.

LEITE, Yonne; FRANCHETTO, Bruna. 500 anos de línguas indígenas no Brasil. In: CARDOSO, Suzana A. M.; MOTA, Jacyra; MATTOS E SILVA, Rosa V. (orgs.), 500 anos de línguas indígenas no Brasil. Salvador, BA: Secretaria da Cultura e Turismo do Estado da Bahia, 2006, p. 15-62.

LOUREIRO-PORTO, Lucía; SAN MIGUEL, Max. Language choice in a multilingual society. A view from complexity science. In: MUFWENE, Salikoko S.; COUPÉ, Christophe; PELLEGRINO, François (eds.). *Complexity in Language: Developmental and Evolutionary Perspectives*. Cambridge/New York: Cambridge University Press, 2017. p. 187-217.

LÜPKE, Friederike. Uncovering small-scale multilingualism. *Critical Multilingualism Studies*, vol. 4, n. 2, p. 37–74, 2016.

MAIA, Moisés (Akito); MAIA, Tiago (Kimâro). *Isâ yekisimia masîke: O conhecimento de nossos antepassados*. Uma narrativa oyé. Narradores indígenas do Rio Negro, vol. 6. São Gabriel da Cachoeira: COIDI/FOIRN, 2004.

MUFWENE, Salikoko S. Language endangerment: What have pride and prejudice got to do with it? In: JOSEPH, Brian; DESTAFANO, Johanna; JACOBS, Neil G.; LEHISTE, Ilse (eds.). *When languages collide*. Perspectives on language conflict, language competition, and language coexistence. Columbus: Ohio State University Press, 2003. p.324-346.

MUFWENE, Salikoko S. Language birth and death. *Annual Review of Anthropology*, Stanford, vol. 33, n. 1, p. 201–222, 2004.

MUFWENE, Salikoko S. *Language evolution*. Contact, competition and change. London/New

York: Continuum International Publishing Group, 2008.

MUFWENE, Salikoko S. The emergence of complexity in language. An evolutionary perspective. In: MASSIP-BONET, Àngels; BASTARDAS-BOADA, Albert (eds.), *Complexity perspectives on language, communication and society*. Dordrecht: Springer, 2013. p.197-218.

MUFWENE, Salikoko S. Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language*, Ann Arbor, vol. 93, n. 4, p.e202–e223, 2017.

MUFWENE, Salikoko S.; COUPÉ, Christophe; PELLEGRINO, François. *Complexity in Language*. Developmental and evolutionary perspectives. Cambridge: Cambridge University Press, 2017.

NEVES, Eduardo G. *Paths in the dark waters: archaeology as indigenous history in the Upper Rio Negro basin, Northwestern Amazon*. 1998. Tese de Doutorado. Indiana University.

NIMUENDAJÚ, Curt. Reconhecimento dos rios Içana, Ayaré e Uaupés. Relatório apresentado ao Serviço de Proteção dos Índios do Amazonas e Acre, 1927. *Journal de La Société Des Americanistes de Paris*, Paris, vol. 39, p. 125–183, 1950.

RODRIGUES, Aryon D. Línguas indígenas: 500 anos de descobertas e perdas. *D.E.L.T.A*, vol. 9, n.1, p. 83-103, 1993.

SAUSSURE, Ferdinand de. *Curso de linguística geral*. São Paulo: Cultrix, 1969.

SHULIST, Sarah. *Transforming indigeneity*. Urbanization and language revitalization in the Brazilian Amazon. Toronto/Buffalo/London: University of Toronto Press, 2018.

SILVA, Wilson de L. Multilingual interactions and code-mixing in northwest Amazonia. *International Journal of American Linguistics*, vol. 86, n.1, p. 133-154, 2020.

STENZEL, Kristine; KHOO, Velda. Linguistic hybridity: A case study in the Kotiria community. *Critical Multilingualism Studies*, vol. 4, n. 2, p. 75–110, 2016.

TRUDGILL, Peter. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press, 2011.

VIOTTI, E.C. Mudança linguística. In: FIORIN, José Luiz (ed.), *Linguística? Que é isso?* São Paulo: Editora Contexto, 2013. p. 137-179.

VIOTTI, E.C. Diário de campo. São Gabriel da Cachoeira, julho de 2016 e julho de 2017. São Paulo, 2017.

ESTEREÓTIPOS NA CONCORDÂNCIA DE GÊNERO EM PROFISSÕES: EFEITOS DE FREQUÊNCIA E SALIÊNCIA

STEREOTYPES IN GENDER AGREEMENT IN PROFESSIONS: EFFECTS OF FREQUENCY AND SALIENCE

Bruno Felipe Marques Pinheiro¹

Raquel Meister Ko. Freitag²

RESUMO

A marcação de gênero no português pode ser binária (“o aluno”/“a aluna” – “o estudante”/“a estudante”) ou por processo de concordância, como em profissões de nome comum de dois gêneros (“o falante”/“a falante”). A distinção da marcação de gênero nestes nomes é resultado das experiências dos falantes com estereótipos compartilhados: na falta de informação gramatical de gênero nos nomes, os falantes constroem representações mentais a partir dos estereótipos de gênero, que sofrem efeitos da frequência e da saliência, parâmetros que também afetam a gramática. A fim de identificar a relação entre os estereótipos e a marcação de gênero, um estudo experimental de julgamento foi desenvolvido considerando o efeito da masculinidade e da feminilidade de nomes comuns de dois gêneros de profissões no português brasileiro. Os resultados apontam que, embora sejam nomes comuns de dois gêneros, o modo como uma profissão foi apresentada ao participante interferiu no resultado de julgamento, denotando efeito de saliência. A frequência prototípica da profissão interferiu positivamente no julgamento. Concluímos que os parâmetros da estereotipia de gênero, medidos pela saliência e frequência, interferem no julgamento das profissões quanto ao gênero nos nomes comuns de dois gêneros.

Palavras-chave: Gênero gramatical; Estereótipos; Saliência; Frequência.

ABSTRACT

Gender marking in Portuguese can be binary (‘o aluno’/‘a aluna’ – ‘the.MASC student’/‘the.FEM student’) or by syntactic process of agreement, like in profession naming by common gender nouns (‘o falante’/ ‘a falante’ – ‘the.MASC speaker’/‘the.FEM speak’). The distinction of gender marking on these nouns results from the speakers’ experiences with shared stereotypes: due to the lack of grammatical information of gender in these names, the speakers create mental representations based on stereotypes that are affected by frequency and salience, features that also affect grammar. To identify the relationship between stereotypes and gender marking, an experimental study of judgment was carried out considering the effect of masculinity and femininity on the common gender nouns in Brazilian Portuguese. The results indicate that, although being common gender nouns, the manner in which the profession was presented to the participants interfered with the result, revealing the salience effect. The prototypical frequency of the profession interfered positively in the rating. In conclusion, the characteristics of gender stereotyping, as measured by salience and frequency, interfered in the rating of profession naming in terms of their gender for common gender nouns.

Keywords: Grammatical gender; Stereotypes; Salience; Frequency.

1 Mestrando em Estudos Linguísticos pelo Programa de Pós-graduação em Letras da Universidade Federal de Sergipe. E-mail: bpinnheiro@hotmail.com

2 Doutora em Linguística pela Universidade Federal de Santa Catarina. Professora do Departamento de Letras Vernáculas da Universidade Federal de Sergipe. E-mail: rkofreitag@uol.com.br

1 Introdução³

Embora a Constituição Federal de 1988, em seu artigo 5º, preconize que “Todos são iguais perante a lei, sem distinção de qualquer natureza”, foi somente a partir de 3 de abril de 2012, com a promulgação da Lei Federal n. 12.605, que as mulheres puderam ter seus diplomas e certificados flexionados com o gênero gramatical correspondente ao seu sexo biológico ou identidade social para designar profissão e o grau obtido. Até então, os diplomas eram todos flexionados no gênero masculino: licenciado em Letras, seja Raquel ou Bruno.

Como regra geral no português, a flexão do gênero feminino ocorre pela oposição ao morfema zero, neste caso, o gênero masculino, e a desinência *-a* no feminino é acrescentada aos nomes flexionados (CÂMARA Jr., 1975). A flexão do gênero gramatical é caracterizada por dois aspectos: i) é marcada de forma binária (masculino e feminino), não havendo, de início, gênero neutro; ii) e pelo processo morfológico de concordância, mesmo não havendo de início marca explícita de gênero, como em “estudante”, ao se relacionar com outros termos, emergirá na sentença marca de gênero: “estudante muito dedicado” ou “estudante muito dedicada” (SCHWINDT, 2018, p.746).

A distinção da marcação de gênero nos nomes/substantivos comuns de dois gêneros também pode ser considerada como resultado das experiências com os estereótipos compartilhados entre os falantes para a marcação do gênero. Por exemplo, no inglês estadunidense, pesquisas sugerem a relação entre a marcação de concordância de gênero e estereótipos: na falta de informação gramatical de gênero nos nomes, as pessoas escolhem um gênero a partir de representações mentais construídas por estereótipos de gêneros sociais (GARNHAM, DOEHREN, GYGAX, 2015; GYGAX, GARNHAM, DOEHREN, 2016).

Estereótipos são representações mentais que determinados grupos sociais constroem a partir de traços e características comportamentais amplamente compartilhados entre os membros dos grupos sociais sobre determinado objeto na sociedade (HEWSTONE, FINCHAM, FOSTER, 2005). Assim, o estereótipo é compreendido enquanto “imagem mental hiperssimplificada de uma determinada categoria (normalmente) de indivíduo, instituição ou acontecimento, compartilhada em aspectos essenciais, por grande número de pessoas” (TAJFEL, 1981, p. 161). Os estereótipos não são, necessariamente, acompanhados de preconceito, ou seja, uma predisposição favorável ou

³ Agradecemos aos dois pareceristas anônimos que pontuaram aspectos que mudaram a direção da análise e aprimoramento da visualização dos resultados, contribuindo substancialmente para o aprimoramento do manuscrito. Agradecemos ainda mais às editoras deste dossiê, Marije Soto e Ana Regina Calindro, pela condução colaborativa e generosa do processo de revisão do texto, que foi aprimorado significativamente em função de suas considerações. Erros remanescentes são de nossa responsabilidade.

desfavorável em relação a qualquer membro de grupo social ou categorias, mas atuam na construção de possíveis representações sociais de variados temas postos em circulação na sociedade à medida que, não apenas guiam o comportamento dos indivíduos, como constroem e remodelam o ambiente em que ocorrem os estereótipos (CAMINO; TORRES, 2011, p. 87).

Os efeitos da estereotipia do gênero social podem emergir na concordância de gênero dos nomes na língua, especificamente no que diz respeito à marcação do gênero gramatical. No inglês, evidências de como traços de crenças e estereótipos são transpostos à gramática da língua são identificadas em estudos envolvendo estereotipia de gênero e marcação de gênero implícita (GARNHAM, DOEHREN, GYGAX, 2015; GYGAX, GARNHAM, DOEHREN, 2016). Assim como no inglês, para a marcação no português brasileiro, o efeito da representação mental do estereótipo de gênero social pode ser observado em nomes de profissões comuns de dois gêneros, como, “o/a dentista”, “o/a soldado”, “o/a babá”, dentre outros, para os quais, possivelmente há interferências no processo de concordância no nível sintático com o determinante. A fim de contribuir para a identificação da relação entre os estereótipos e a marcação de gênero no português brasileiro, medimos a influência dos estereótipos de gênero no julgamento dos falantes quanto a uma escala de masculinidade e de feminilidade⁴ na marcação dos nomes comuns de dois gêneros relacionados às profissões.

A hipótese é de que, não havendo informação de gênero gramatical nos nomes, as pessoas passam a marcar a concordância de gênero baseadas nas representações mentais de gênero (informação de gênero social) dos estereótipos na estrutura linguística. Para integrar a estereotipia à gramática, adotamos uma abordagem sócio-cognitiva, considerando os efeitos da saliência e da frequência na marcação do gênero gramatical, a partir da exposição a padrões de marcação de gênero gramatical e de estereótipos. Para tanto, tomamos como referência um levantamento da frequência de marcação de gênero em nomes de profissões comuns de dois gêneros na internet no paradigma *Google as a corpus* (ROBB, 2003; DIEMER, 2008; SHA, 2010) e comparamos com o julgamento atribuído a estes mesmos nomes de profissões, em função do modo como são apresentadas as tarefas (julgamento de masculinidade e de feminilidade da mesma profissão) e em função do sexo/gênero do participante, tendo como objetivo avaliar pistas do efeito de saliência: se o *priming* da profissão e o gênero de frequência prototípica ativam estereótipos e interferem na nota de julgamento.

4 Os termos “feminilidade” e “masculinidade” dizem respeito às representações de “ser mulher” e “ser homem” relacionadas aos padrões de comportamentos dos indivíduos, estabelecidos *a priori*. Ou seja, refere-se à representação mental de sexo/gênero que ocorre desde o nascimento (o bebê não tem liberdade de escolha) até o momento que os indivíduos precisam se adaptar aos padrões de feminilidade e masculinidade nos grupos sociais (WELZER-LANG, 2001). Destacamos que o objetivo do estudo não é capturar significados sociais quanto à identidade de gênero. Isto é, o objetivo não é adentrar na discussão sobre as categorias sexo/gênero enquanto construção social e seu efeito na representação mental pelo viés da transversalidade, nem discutir o escopo das identidades de gênero (BUTLER, 1990).

2 Estereotipia, Frequência e Saliência

A frequência de uma realização, como, por exemplo, o nome de uma profissão comum de dois gêneros e um artigo definindo o gênero do referente no mundo pode ser rotinizada e automatizada. Assim, o que eram dois pacotes informacionais (a profissão e o gênero) passam a ser processados como uma única unidade. Este é o efeito da autonomia, resultado do processo em que realizações morfológicamente complexas de alta frequência perdem sua estrutura interna à medida que se tornam autônomas (BYBEE; HOPPER, 2001; BYBEE, 2006).

Assim como a automatização, o *priming* refere-se ao status de ativação de elementos linguísticos na memória de trabalho: elementos linguísticos ativados se fazem presentes, não só no nível lexical, mas também no nível gramatical. O processamento não é influenciado apenas por *priming*, mas também pelo conhecimento linguístico de longo prazo do falante. A co-ocorrência frequente de padrões facilita o processamento, e as escolhas linguísticas dos falantes são previsíveis a partir das experiências individuais (DIESSEL; HILPERT, 2016).

Em uma perspectiva cognitivista, ou baseada no uso, a capacidade humana de processar a língua envolve não só o conhecimento linguístico estrito e interno, mas a sua relação com o mundo externo; a frequência e a saliência de um dado elemento interferem no modo como o conhecimento linguístico é armazenado, ativado e processado (BYBEE; HOPPER, 2001; BYBEE, 2006; DIESSEL; HILPERT, 2016, dentre outros).

Nas abordagens baseadas no uso, a gramática está relacionada às capacidades gerais do cérebro humano para categorizar informações e organizá-las por identidade, similaridade e diferenças, que são aplicadas às experiências de língua do falante, que permitem categorizá-las e armazená-las na memória. O conhecimento gramatical é derivado da experiência dos falantes da língua com a frequência de cadeias linguísticas concretas. Esta concepção prevê um componente probabilístico na construção da gramática (BOD; HAY; JANNEDY, 2003).

A automatização é o processo psicológico que dirige o conhecimento do falante da língua e o processamento de padrões de coocorrência; a qual é impulsionada pela frequência de ocorrência e não é restrita à língua, envolvendo processamento de conexões associativas entre conceitos e características de categoria em geral (FISK; SCHNEIDER, 1984). Assim, a frequência de experiências com a língua contribui para a rotinização, convencionalização e automatização no processo de categorização e armazenamento. Bybee (2006, p. 772) destaca que a atuação da frequência pode ter efeito de redução, de conservação ou de autonomia na gramática; assim, a frequência afeta a organização e o

desenvolvimento da estrutura morfológica e sintática (BAAYEN, 2003; DIESSEL; HILPERT, 2016).

O processamento de todas as experiências, no entanto, é cognitivamente custoso e dispendioso. Nem tudo que é frequente é necessariamente importante em termos de demanda de atenção para que seja processado; para isso, construímos atalhos cognitivos que filtram o que é importante para ser processado: filtros de saliência. Saliência pode ser definida como um critério de demanda de atenção, que varia em função das experiências. Embora se pressuponha que o que é pouco comum seja mais saliente, simplesmente pode não o ser, por exemplo, quando não está no campo de prioridade de processamento (GIBSON; SCHÜTZEL; SALOMON, 1996); filtros de saliência podem ser sensíveis a estímulos raros ou pouco frequentes, ou por estímulos com importância para a aprendizagem.

Diferentes abordagens tentam explicar o que é saliente. Pelo viés sócio-cognitivo, Blumenthal-Dramé, Hanulíková e Kortmann (2017, p.4) propõem duas direções de saliência, *top-down* e *bottom up*. Um estímulo pode ser saliente porque é cognitivamente pré-ativado (saliência *top-down*). Este tipo de saliência pode ocorrer se um estímulo é esperado, eleito como parte de uma rotina cognitiva, se foi recentemente mencionado, ou devido às intenções atuais do falante. É o caso da saliência que ocorre no nível semântico-pragmático, com preferências na interpretação de enunciados figurativos, antecedentes pronominais, implicaturas, e *links* discursivos (GEERAERTS, 2000; GIORA, 2003; JASZCZOLT; ALLAN, 2011).

Do ponto de vista cognitivo, os efeitos de superfície que são atribuídos à saliência são decorrentes da interação entre frequência, familiaridade e convencionalidade que levariam à prototipia e estereotipia (GIORA, 2003, p. 17). A abordagem sócio-cognitiva pressupõe que os falantes não são apenas limitados pelas condições da sociedade, os valores pactuados do que é aceito e não aceito em cada época e em cada cultura, mas que eles também moldam essas condições. Na interação, processos de cooperação e egocentrismo estão presentes em maior ou menor medida. Enquanto a cooperação é uma prática direcionada à intenção, medida por relevância, o egocentrismo é um traço orientado para a atenção medido pela saliência. (KECSKES, 2017). Egocentrismo significa que os interlocutores ativam a informação mais saliente para a sua atenção na construção (falante) e compreensão (ouvinte) dos enunciados. O processamento linguístico implica que o que é saliente ou acessível ao falante também estará acessível aos seus interlocutores (GIORA, 2003; BARR; KEYSAR, 2005; KECSKES, 2017).

Na interação, o falante usará os recursos linguísticos (por exemplo, unidades lexicais, estruturas sintáticas) que julga mais relevantes para expressar suas intenções e/ou objetivos comunicativos. Na

abordagem sócio-cognitiva, tanto os falantes quanto os ouvintes confiam em seu conhecimento mais acessível e saliente expresso no processamento linguístico. Por causa de sua base de conhecimento diferente, a frequência/rotinização de seu conhecimento na situação e os recursos atencionais disponíveis para processar os itens salientes, o conhecimento dos falantes tem diferentes níveis de saliência.

Schmid e Günther (2016, p.1) propõem um tratamento unificado para a saliência considerando o ponto de vista do falante, a motivação da saliência, o mecanismo e o contexto de ativação. A saliência é baseada em conhecimento lexical (palavras) e em contexto linguístico (aquilo que já foi dito antes), contexto situacional (participantes, tempo, lugar, configuração), contexto social (tipo de evento social, papéis sociais e relações entre os participantes) e contexto cognitivo geral (conhecimento geral e conhecimento linguístico armazenado na memória de longo prazo). Nesta abordagem, saliência e frequência interagem, o que a torna apropriada aos propósitos desta investigação: a forma como uma profissão é apresentada, se masculina ou se feminina, pode atuar como um gatilho de saliência para a avaliação do falante (KENNISON; TROFE, 2003; ARSENIIEV-KOEHLER, 2019).

Um tipo de experiência decorre da frequência a que cada falante é exposto a cada profissão no seu pareamento com o gênero. Não temos acesso a todas as experiências dos falantes, mas podemos medir a frequência de recorrência da associação de uma profissão a um gênero gramatical, e como esta frequência afeta o julgamento de profissões como mais femininas ou menos femininas.

3 Estereótipos e marcação de gênero

De acordo com Corbett (2007, p. 749), em um universo de 256 línguas, mais da metade não tenha sistema de marcação de gênero, muitas línguas dispõem de sistema com dois valores, três, quatro ou até mesmo existem línguas com cinco ou mais valores (por exemplo, as línguas bantu). Logo, tais possibilidades estão associadas à semântica do gênero gramatical. No português brasileiro, o gênero é tratado como categoria morfológica de flexão (CÂMARA Jr., 1970, p.74), mas a realização dos contrastes de gênero “não pode ser atribuída à flexão, mas se relacionam com a constatação de que a variação em gênero não é obrigatória, nem sistemática” (VILLAVA, 2000, p. 219, grifo nosso).

O gênero, então, passa a ser tratado a partir de dois aspectos: a sua não obrigatoriedade e assistemática. No que se refere a não obrigatoriedade, a marcação de gênero nos nomes ocorre a partir de uma variação semântica e lexicalmente condicionada, “livro”, “casa”; “cônjuge”; “testemunha” vs. “aluno/aluna”; “duque/duquesa”; “águia-macho/águia-fêmea”. Quanto à assistemática, a marcação de gênero dispõe de maiores possibilidades nos nomes de construções: palavras que partilham do mesmo radical, mas pertencem a classes temáticas diferentes (“menino”/“menina”); o contraste de

gênero pode ser realizado por derivação (“conde”/“condessa”), ou por composição (“águia-macho”/“águia-fêmea”), ou por pares não relacionados morfologicamente (“homem”/“mulher”) (VILLALVA, 2000, p.219).

O gênero pode adquirir um caráter de conteúdo referencial: os nomes no português são dotados de informação morfológica de gênero, assim como as línguas de matriz indo-europeia, como o espanhol e italiano, mas também adquirem oposições semânticas entre os seres inanimados e animados, repartidos em dois sexos (masculino e feminino): “No Português, os nomes requerem uma especificação, tanto no que diz respeito ao valor de gênero que lhes está associado, como à possibilidade de participarem em contrastes de gênero” (VILLALVA, 2000, p. 242).

A possibilidade de haver contrastes de gênero nos nomes, então, é semanticamente relacionada com a oposição [+/- animado], acarretando inclusive na limitação dos nomes [+ animado], em que o contraste de gênero gramatical coincide geralmente com o contraste semântico de gênero. Com isso, os nomes comuns de dois gêneros podem evidenciar esses contrastes de gênero, implicando, inclusive, nas realizações formais. É comum a produção de nomes como: “mestre/mestra”, “monge/monga”, “parente/parenta”, “presidente/presidenta”.

A marcação de gênero também esbarra na problemática de ser marcado e não marcado na língua: muitas vezes, o gênero gramatical não está associado intimamente ao sexo biológico dos seres. Por exemplo, na norma padrão, “testemunha” sempre é feminino, seja para se referir a homem e mulher, ao mesmo modo que “cônjuge”, sempre é masculino, seja para se referir ao esposo ou à esposa.⁵ O gênero, na realidade, é uma distribuição das classes mórficas para os nomes da mesma forma para conjugações verbais. A diferença é que a oposição masculino/feminino serve para distinguir os seres com certas qualidades semânticas e para os animais e pessoas servem para distinguir o sexo (por exemplo, menino/menina ou urso/ursa) em que, do ponto de vista morfossemântico, o masculino é a forma geral, não marcada, enquanto o feminino indica a forma marcada na língua (CÂMARA Jr., 1970, p. 78).

No entanto, na tradição gramatical, há incoerência e confusão na descrição do gênero em português. Por exemplo, para a gramática normativa, “mulher” é “feminino” de “homem”. “O substantivo ‘mulher’ sempre é feminino, ao passo que outro substantivo, a ele semanticamente relacionado, é sempre do gênero masculino” (CÂMARA Jr., 1970, p. 79).

A oposição entre a informação morfológica de gênero e oposições semânticas pode se relacionar à estereotipia porque o gênero pode adquirir caráter de conteúdo referencial. Com isso, a partir das construções estereotipadas enquanto esquemas sociais, as pessoas processam informações de

⁵ É assim que todos os dicionários de língua portuguesa apresentam o gênero para o item. No entanto, o uso da forma com síncope na penúltima sílaba e flexionada no feminino é atestado: “A possibilidade, por exemplo, de uma mulher, uma conje, seja morta pelo seu conje”. <https://tv.uol/17mjX>

percepção e cognição, possibilitando a construção de mecanismos para a compreensão do mundo complexo no qual vivem (AUGOUSTINOS; WALKER, 1995, p.33). Uma vez que os nomes comuns de dois gêneros não apresentam uma informação gramatical morfológicamente explícita, para realizarem o processo de marcação desses nomes os falantes recorrem, na maioria das vezes, a estereótipos construídos e compartilhados pelos grupos sociais aos quais pertencem.

Os estereótipos são generalizações a partir do processo cognitivo da categorização, com a simplificação ou sistematização de apreensões complexas na sociedade com vista à adaptação cognitiva e comportamental da complexidade que existe de informação no meio social. Os estereótipos passam a ser sociais no momento em que são compartilhados por um grande número de pessoas no interior de grupos sociais ou entidades (TAFJEL, 1981, p.162).

Os estereótipos baseiam-se em dois pressupostos: o primeiro corresponde à análise de processos individuais, seja representacional ou motivacional, para a compreensão de comportamento e interação sociais; já o segundo é consequência do primeiro, corresponde às possibilidades de situações decorrentes da interação entre o comportamento e o seu contexto social, isto é, os estereótipos ativam e selecionam mecanismos individuais ou modos pré-existentes para simplificar modos complexos. Os estereótipos, ao serem compartilhados por um grande número de pessoas, são estruturados pelas relações entre os grupos sociais em larga escala (TAFJEL, 1981, p. 163).

No inglês, experimentos psicolinguísticos evidenciam que, quando não há informação gramatical de gênero nos nomes, estudantes universitários, ao ler frases com substantivos comuns de dois gêneros, baseiam-se em suas representações mentais de gênero nos estereótipos. Ou seja, a informação estereotipada interage com o gênero gramatical para produzir comportamentos mais complexos (OAKHILL et al., 2005; GYGAX, GARNHAM, DOEHREN, 2016).

A verificação do efeito da correspondência do estereótipo com informações de gênero requer um desenho de experimento em que as frases são elaboradas de maneira que a informação de gênero está explicitada em construções gramaticais relacionadas ao vestuário ou características relacionadas ao sexo biológico, e não nos nomes ou pronomes, para não ter marca morfológica explícita. O estudo de Garnhaw, Oakhill e Reynolds (2002, p. 26), no inglês, para medir marcação de gênero e influência dos estereótipos aponta para efeitos no julgamento e no tempo de resposta: o julgamento é mais positivo para as frases que carregam características biológicas (por exemplo, informações a respeito de homens terem barba e só as mulheres terem bebês), enquanto o efeito de incompatibilidade do gênero é negativo para as frases que carregam características relacionadas às roupas. Já em relação ao tempo gasto, as frases ligadas ao vestuário demandam maior tempo (os participantes demandam mais tempo de processamento), enquanto as frases com características biológicas demandam menor tempo (os participantes reconhecem a informação de gênero mais rápido).

Os estereótipos de gênero podem influenciar na resolução de correferência entre nomes e pronomes em palavras específicas: em experimentos de compreensão de leitura de frases com substantivos fortemente marcados com estereotipia masculina e substantivos fortemente marcados com estereotipia feminina. Por exemplo, em um teste, os participantes precisavam realizar a leitura de frases embaralhadas com sujeito gramatical com uma palavra fortemente estereotipada para se referir a homens (“eletricista”) ou mulheres (“babá”), sendo o sujeito retomado pelos pronomes *he* ou *she*. Os resultados apontam que o tempo de leitura das frases em que o gênero do pronome não corresponde ao estereótipo da palavra do pronome antecedente é significativamente maior do que o tempo de leitura das frases em que os gêneros dos pronomes e antecedentes coincidem (KENNISON; TROFE, 2003).

Em títulos de profissões retirados de artigos do jornal *The New York Times*, nos anos de 2002-2007, de 237 títulos para profissões, 221 tinham conotação mais masculina, e 69 tinham conotação mais feminina. Por meio de análise de similitude, as palavras foram compartilhadas pelos significados utilizando vetores, desenvolvendo uma escala de conotação de gênero que variou entre -1 para conotação masculina mais forte e 1 para conotação feminina mais forte. Os resultados obtidos sugerem quatro tipos de profissões: “presidente” (*chairperson*), “vendedor” (*salesperson*), “empresário” (*businessperson*) e “policia” (*police officer*). As três primeiras profissões, não marcadas no inglês, indicaram ser julgadas mais femininas do que masculinas. Todavia, a forma não marcada “policia” (*police officer*) permanece fortemente associada à masculinidade, muito mais do que a forma marcada “policia homem” (*policeman*), o que sugere efeito mais forte nas conotações de feminilidade do que masculinidade na estrutura linguística (ARSENIEV-KOEHLER, 2019).

Estas evidências sugerem diferenças na forma como homens e mulheres reagem à ativação do estereótipo, a partir da frequência de execução das tarefas desempenhadas por homens e mulheres, afetando o reconhecimento da estrutura morfológica. Nessa perspectiva, ao julgar frases ou palavras sem marca explícita de gênero com informações de estereótipos, a manipulação de variáveis morfológicas e as pistas de comportamentos sociais permitem ampliar a compreensão do processamento no nível morfossintático, permitindo a integração entre a informação social e modelos cognitivos para o processamento e julgamento linguístico (FREITAG, 2020).

4 Método

A fim de avaliar o efeito de saliência, se o *priming* da profissão e o gênero de frequência prototípica ativam estereótipos e interferem na nota de julgamento, desenvolvemos um estudo experimental em que, inicialmente, foram selecionadas 30 palavras comuns de dois gêneros relacionadas às profissões

retiradas de abonações em gramáticas do português, sem considerar o controle de contextos linguísticos (como número de sílaba, complexidade morfológica, base).

Nas abonações gramaticais, não há indicação do gênero mais frequente, ou prototípico, associado à profissão com nome comum de dois gêneros. Como a frequência é um parâmetro cognitivo que, por hipótese, atua na construção do estereótipo, foi preciso verificar a frequência da ocorrência dos nomes de profissões em função dos gêneros. Inicialmente, tentamos estabelecer estes parâmetros a partir de *corpora* de fala, mas o tipo do método de coleta de dados não favorece os nomes em questão. Os *corpora* escritos já compilados também não foram adequados por não representarem o estado de língua atual.

Para estabelecer os valores de referência das profissões quanto à frequência por gênero, seguimos uma abordagem conhecida como *Google as a corpus* (ROBB, 2003; DIEMER, 2008; SHA, 2010), utilizada para o aprendizado de máquina quanto a padrões linguísticos, mas também para tradução. A escolha dessa abordagem se sustenta na atualidade, diversidade de gêneros e tipos textuais que são acionados no motor de busca do Google, em termos de frequência.

4.1 Design

Após o levantamento de palavras comuns de dois gêneros relacionadas ao tópico “profissão” em gramáticas contemporâneas do português, foram selecionadas 30 palavras-estímulos para compor um teste de julgamento de feminilidade e de masculinidade: “acrobata”, “camelô”, “chofer”, “comandante de navio”, “dj”, “eletricista”, “frentista”, “intérprete de libras”, “servente”, “taxista”, “gari”, “agente de trânsito”, “assistente social”, “atendente”, “atleta”, “babá”, “dentista”, “estilista”, “estudante”, “fiscal de obra”, “gerente de agência bancária”, “jornalista”, “jurista”, “lojista”, “motorista”, “piadista”, “pianista”, “policial”, “repórter” e “soldado”.

A tarefa experimental foi criada no programa computacional *Psychopy* (PEIRCE, 2007), envolvendo uma bateria de treino (2 itens) e uma bateria de teste com o julgamento em uma escala de 1 a 7 do grau de feminilidade ou de masculinidade de cada uma das 30 profissões selecionadas. O número 1 correspondia ao menor grau de feminilidade ou masculinidade até o número 7 correspondendo ao maior grau de feminilidade ou masculinidade. A apresentação do teste foi em tela preta e letras brancas com o intervalo de tempo correspondente a cinco segundos. Considerando o efeito de *priming*, a aplicação do teste ocorreu em duas etapas diferentes, ativadas por dois comandos (“O quão masculina é essa profissão?” ou “O quão feminina é essa profissão?”), em dois grupos diferentes de participantes voluntários, universitários de diferentes cursos, equilibrados quanto ao

gênero autoindicado, residentes na capital Aracaju e no interior do estado de Sergipe. O primeiro grupo (n = 50, dos quais 25 homens e 25 mulheres) foi convidado a julgar o grau de masculinidade de cada uma das profissões; o segundo grupo (n = 50, dos quais 25 homens e 25 mulheres) foi convidado a julgar o grau de feminilidade. O comando da bateria de testes com o julgamento apresentado foi o seguinte: (i) o quão masculina esta profissão é? (ii) o quão feminina esta profissão é? Após cada comando do teste, era apresentada a palavra correspondente à profissão selecionada de maneira aleatória pelo programa.

4.2 Constituição da amostra

A coleta de dados para o teste experimental aconteceu na Universidade Federal de Sergipe, no campus Prof. Aloísio de Campos, localizado na cidade de Cristóvão/Sergipe, e contou com a colaboração de 100 participantes, estudantes universitários de diferentes cursos do campus. Na tarefa experimental, a escolha dos estudantes ocorreu de forma voluntária entre os intervalos das aulas. A aplicação do teste aconteceu no Laboratório Multiusuário de Informática e Documentação Linguística (LAMID), em ambiente controlado quanto aos estímulos paralelos (cabine). Os participantes assinaram termo de consentimento livre e esclarecido (CAAE: 0386.0.107.000-11) e forneceram as informações de autoidentificação de gênero.

4.3. Valor de referência para a frequência das profissões

Para ter um valor de medida de frequência das profissões, partindo do paradigma *Google as a corpus* (ROBB, 2003; DIEMER, 2008; SHA, 2010), procedeu-se à coleta das frequências de ocorrência de cada um dos 30 itens do teste em <https://www.google.com.br>, em páginas em português e a busca pela expressão “ao pé da letra” (para diferenciar “babá” de “baba” e “camelô” de “de camelo”), opção dada pelo buscador. A frequência de cada profissão foi computada quanto à frequência global do item (Tabela 1), e às frequências como profissão feminina e como profissão masculina. Para minimizar imprecisões, já que a busca foi por colocações, sem anotação morfossintática, os parâmetros da pesquisa foram o nome da profissão antecedido de “uma” (frequência feminina) e o nome da profissão antecedido de “um” (frequência masculina). O artigo definido não foi considerado na busca porque no feminino “a” poderia se sobrepor à preposição “a”; também não foram considerados outros determinantes, quantificadores ou numerais, por não permitirem um controle do equilíbrio.

Os valores obtidos para as profissões foram considerados como uma variável independente nominal (frequência prototípica), com a conversão das frequências em função da recorrência quanto ao gênero em escala (baixa: menos de 20.000 ocorrências; média = entre 20.000 e 100.000 ocorrências; alta = entre 100.000 e 1.000.0000 ocorrências e altíssima = mais de um milhão de ocorrências, para

cada gênero). Por exemplo, a profissão “servente” tem frequência média no gênero feminino, e frequência alta no gênero masculino.

Tabela 1 - Frequência das profissões pelo número de ocorrências nos resultados do Google.

Distribuição das profissões								
<i>Critério de frequência no Google</i>								
Profissões	ocorrências (feminino)	%	ocorrências (masculino)	%o	total	range	frequência (feminino)	frequência (masculino)
Acrobata	10.600	28,3	26.800	71,7	37.400	43,3	baixa	média
Agente de trânsito	57.500	15,4	317.000	84,6	374.500	69,3	média	alta
Assistente social	709.000	79,6	182.000	20,4	891.000	59,1	alta	alta
Atendente	1.080.000	51,7	1.010.000	48,3	2.090.000	3,3	altíssima	altíssima
Atleta	1.390.000	20,3	5.450.000	79,7	6.840.000	59,4	altíssima	altíssima
Babá	1.500.000	94,0	95.100	6,0	1.595.100	88,1	altíssima	média
Camelô	3.170	5,8	51.300	94,2	54.470	88,4	baixa	média
Chofer	6.190	13,0	41.300	87,0	47.490	73,9	baixa	média
Comandante do navio	3.530	3,5	96.100	96,5	99.630	92,9	baixa	média
Dentista	231.000	15,4	127.0000	84,6	1.501.000	69,2	alta	altíssima
Dj	106.000	3,8	272.0000	96,2	2.826.000	92,5	alta	altíssima
Eletricista	35.200	12,7	241.000	87,3	276.200	74,5	média	alta
Estilista	648.000	47,0	730.000	53,0	1.378.000	6,0	alta	alta
Estudante	3.210.000	34,3	6.140.000	65,7	9.350.000	31,3	altíssima	altíssima
Fiscal de obra	11.000	10,5	93.300	89,5	104.300	78,9	baixa	média
Frentista	19.600	10,6	166.000	89,4	185.600	78,9	baixa	alta
Gari	24.300	23,7	783.00	76,3	102.600	52,6	média	média
Gerente de banco	6.020	22,0	21.400	78,0	27.420	56,1	baixa	média
Intérprete de libras	13.900	33,9	27.100	66,1	41.000	32,2	baixa	média
Jornalista	9.370.000	36,6	16.200.000	63,4	25.570.000	26,7	altíssima	altíssima
Jurista	82.700	11,1	664.000	88,9	746.700	77,8	média	alta
Lojista	33.900	5,7	564.000	94,3	597.900	88,7	média	alta
Motorista	515.000	5,3	9.210.000	94,7	9.725.000	89,4	alta	altíssima
Piadista	2.890	15,0	16.400	85,0	19.290	70,0	baixa	baixa
Pianista	403.000	27,7	1.050.000	72,3	1.453.000	44,5	alta	altíssima
Policial	475.000	10,2	4.180.000	89,8	4.655.000	79,6	alta	altíssima
Repórter	2.100.000	42,4	2.850.000	57,6	4.950.000	15,2	altíssima	altíssima
Servente	28.200	19,2	119.000	80,8	147.200	61,7	média	alta
Soldado	115.000	4,6	2.410.000	95,4	2.525.000	90,9	alta	altíssima
Taxista	98.500	14,5	581.000	85,5	679.500	71,0	média	alta

Fonte: elaboração própria

4.4 Tratamento dos dados

Os dados foram submetidos a tratamento estatístico descritivo e inferencial. A média geral do julgamento das 30 profissões, pelos 100 participantes, foi de 4,51 ($dp = 1,89$), sendo a maior média para “estudante” ($m = 4,82$, $dp = 2,18$) e a menor para “soldado” ($m = 4,52$, $dp = 2,00$). No procedimento de inspeção dos dados, as condições de normalidade da nota de julgamento (teste Shapiro-Wilk, $p < 0,001$) e de homogeneidade de fatores para frequência de ocorrência (teste Bartlett, $p = 0,030$) não foram satisfeitas. O tratamento inferencial foi realizado por análise bivariada (com teste Wilcoxon entre pares e Kruskal-Wallis entre grupos) com a nota de julgamento como variável dependente, a fim de responder as seguintes questões: o modo como o teste foi apresentado interfere na média de julgamento? (hipótese: por efeito de *priming*, a nota de julgamento da profissão pode ser influenciada pelo modo como o comando do teste apresenta, se a profissão é feminina ou masculina); a frequência do gênero prototípico da profissão (Tabela 1) interfere no julgamento? (hipótese: combinações de gênero mais frequentes podem ser rotinizadas e mais facilmente associadas na memória, influenciando no julgamento); e o gênero do participante interfere no julgamento? (hipótese: a experiência dos estereótipos de gênero vivida por cada participante podem afetar o modo como a profissão é julgada). Para verificar a interação entre os efeitos, um modelo linear misto foi construído considerando os efeitos entre grupos e dentro do grupo.

Os procedimentos de análise foram desenvolvidos na plataforma R (R CORE TEAM, 2020), e o conjunto de dados e o script de análise estão disponíveis em <https://osf.io/qt47u/>.

5 Resultados e discussões

As profissões associadas ao estereótipo feminino e as profissões associadas ao estereótipo masculino revelam como as inferências que os participantes fazem dependem da ocorrência dos itens lexicais com a experiência de mundo.

As informações de estereotipia estão fortemente associadas às palavras estímulos, mesmo não havendo marcação explícita de gênero, nem efeito de *priming*. Muitas vezes, essas informações são recuperadas pela compreensão cultural da função e pelos efeitos de saliência e frequência, e estas informações são armazenadas no nível da representação gramatical, para serem ativadas com regularidade e previsibilidade, como propõe a abordagem sócio-cognitiva. O efeito de saliência sociocognitiva decorre da frequência, familiaridade e convencionalidade, o que permite recuperar mais rápido os significados na memória dos participantes durante o processamento das informações (GIORA, 2003, p. 15-16). O modo como o comando do teste foi apresentado aos participantes, se solicitando o julgamento de feminilidade ou de masculinidade, tem efeito nas médias de julgamento das profissões (Figura 1): a média das notas atribuídas às profissões quando apresentadas como femininas ($m = 4,85$, $dp = 1,79$) é maior do que quando apresentadas como masculinas ($m = 4,18$, $dp = 1,93$), e esta diferença é estatisticamente significativa (teste Wilcoxon $p < 0,001$).

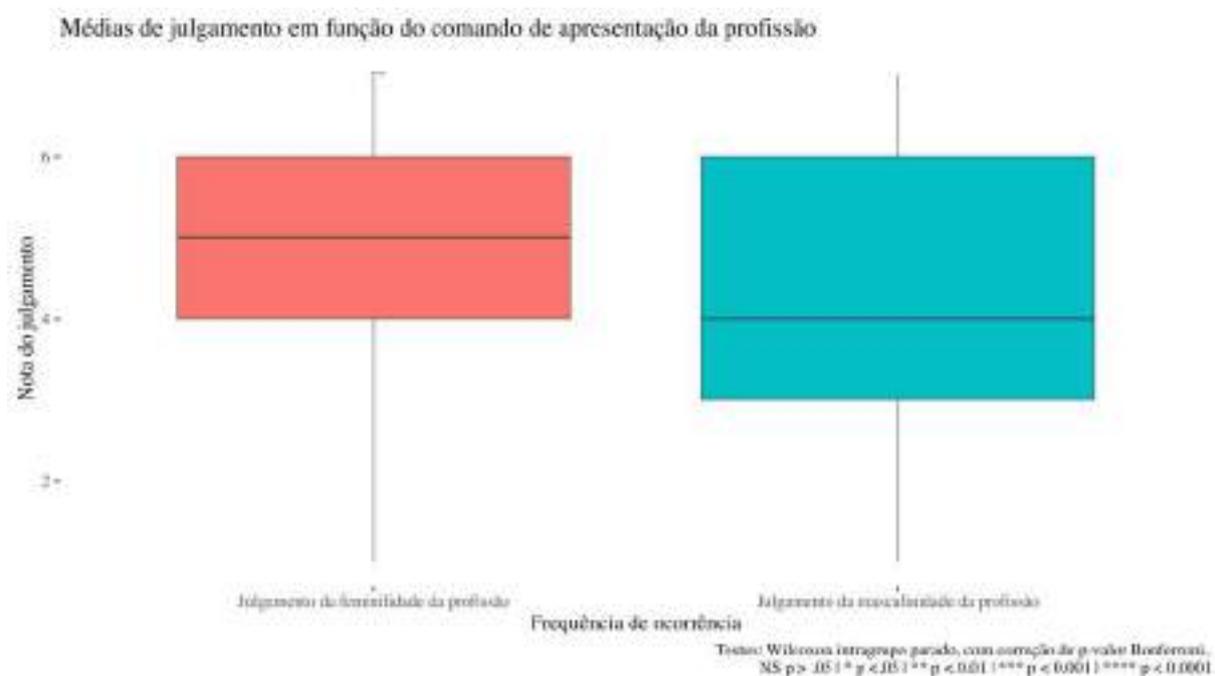


Figura 1 - Distribuição das notas quanto ao comando do teste.

Fonte: elaboração própria.

Na Figura 2, as médias são apresentadas polarizadas em função do comando de apresentação da profissão no teste. Para “camelô”, “dj”, “servente”, “gari”, “piadista”, “agente de trânsito” e “policial”, o modo como o teste foi introduzido não interfere na nota do julgamento, pois as diferenças entre as médias atribuídas às profissões apresentadas como masculinas ou femininas não são estatisticamente significativas (teste Wilcoxon, $p > 0,05$). Neste grupo, somente “piadista” é marcado com vogal temática -a, e é a palavra com recorrência mais baixa dentre todo o conjunto de itens do teste. Neste caso, é possível se considerar o efeito da não saliência, por ser uma palavra pouco frequente, para a qual não há ainda uma associação de estereótipo construída por regularidade e familiaridade.

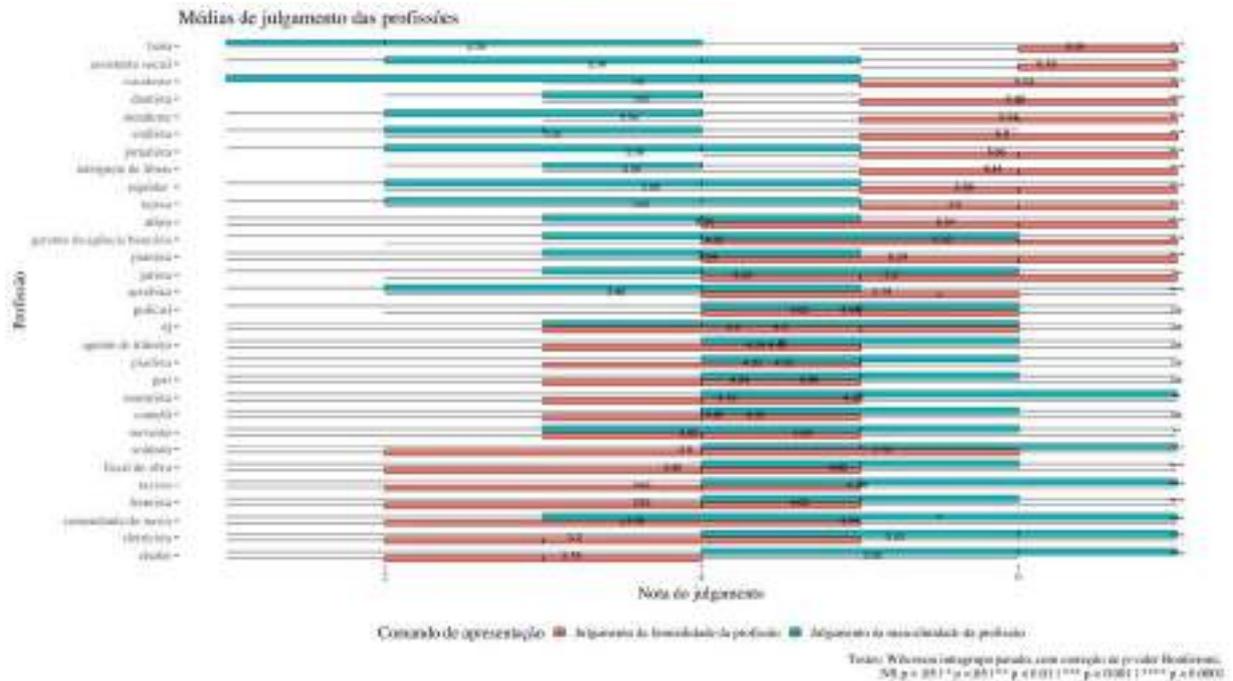
Já “eletricista”, “soldado”, “chofer”, “taxista”, “motorista”, “comandante do navio” e “fiscal de obras” tem nota de julgamento mais alta quando apresentadas como masculinas do que quando apresentadas como femininas (teste Wilcoxon, $p < 0,05$). Já para “pianista”, “atleta”, “repórter”, “lojista”, “dentista”, “estudante”, “jornalista”, “intérprete de libras”, “atendente”, “acrobata”, “assistente social”, “estilista” e “babá”, a nota de julgamento é mais alta quando apresentadas como femininas do que como masculinas (teste Wilcoxon, $p < 0,05$).

Para “camelô”, “dj”, “servente”, “gari”, “piadista”, “agente de trânsito” e “policial”, o modo como o teste foi introduzido não interfere na nota do julgamento, pois as diferenças entre as médias atribuídas às profissões apresentadas como masculinas ou femininas não são estatisticamente significativas (teste Wilcoxon, $p > 0,05$).

A noção de saliência está relacionada à probabilidade de encontrar mulheres e homens executando funções e tarefas relacionadas às profissões, a frequência está correlacionada com a coocorrência de significados no léxico mental dos participantes, “quanto mais frequente o significado, mais rápido de

recuperar” (GIORA, 2003, p. 15)⁶.

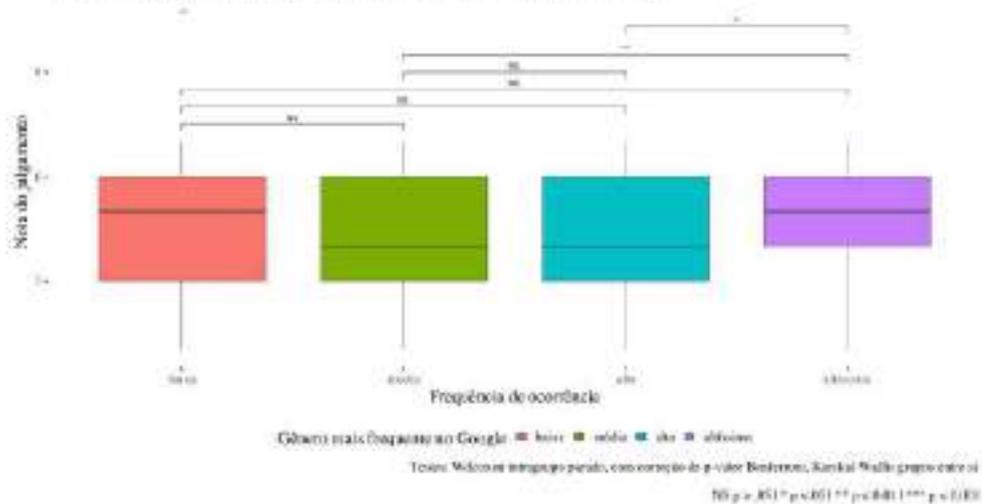
Figura 2 - Diferença entre notas de julgamento para as profissões quanto ao comando.



Fonte: elaboração própria.

A frequência da profissão no Google como masculina está relacionada com a nota de julgamento das profissões (Figura 3): nas profissões de maior frequência com gênero masculino, a nota de julgamento é maior (teste Kruskal-Wallis, $p = 0,003$): média ($m = 4,36$, $dp = 1,79$), alta ($m = 4,43$, $dp = 1,89$) e altíssima ($m = 4,67$, $dp = 1,80$), sendo estatisticamente significativas as comparações pareadas: média – altíssima (teste Wilcoxon $p < 0,001$), alta – altíssima (teste Wilcoxon $p < 0,01$).

Figura 3 - Distribuição das notas quanto à frequência prototípica masculina.
 Médias de julgamento em função da frequência como masculino no Google

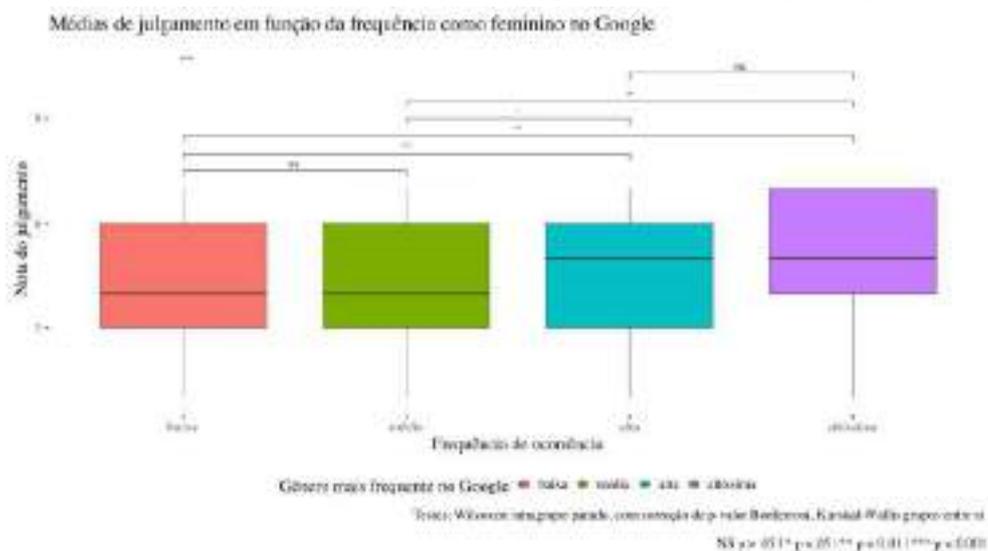


Fonte: elaboração própria.

6 No original: “The more frequent the meaning, the quicker it is to retrieve.” (tradução nossa).

O mesmo efeito se observa com a frequência do gênero da profissão feminina: nas profissões de maior frequência no Google com gênero feminino, a nota de julgamento também é maior: baixa ($m = 4,31$, $dp = 1,85$), média ($m = 4,43$, $dp = 1,79$), alta ($m = 4,61$, $dp = 1,92$) e altíssima ($m = 4,7$, $dp = 2,00$), sendo estatisticamente significativas as comparações entre os grupos baixa – alta (teste Wilcoxon $p < 0,001$), baixa – altíssima (teste Wilcoxon $p < 0,001$), média – alta (teste Wilcoxon, $p = 0,05$) e média – altíssima (teste Wilcoxon, $p > 0,01$).

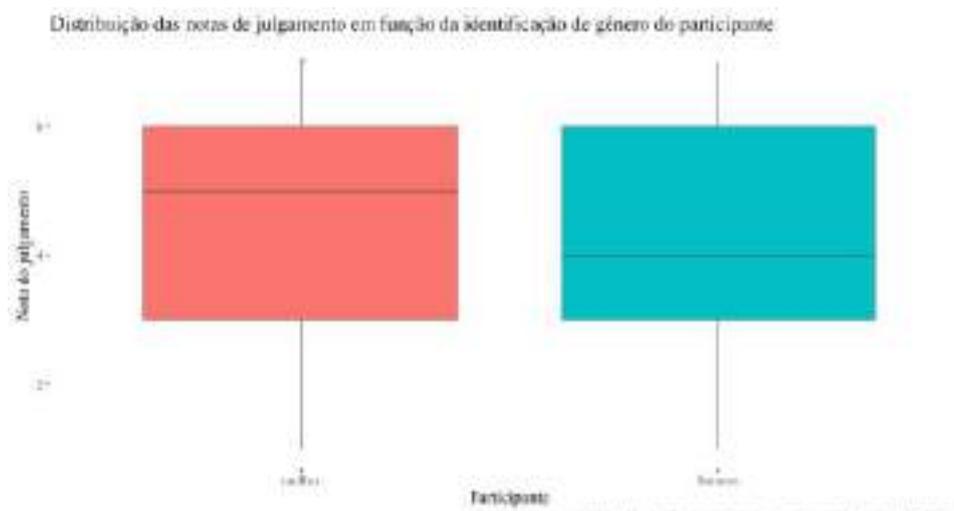
Figura 4 - Distribuição das notas de julgamento quanto à frequência prototípica feminina.



Fonte: Elaboração própria.

O efeito da frequência é estatisticamente significativo tanto quando o gênero é masculino quanto é feminino, o que sugere que o efeito maior é da frequência em si, e não necessariamente do gênero gramatical associado.

Figura 5 - Distribuição das notas de julgamento quanto ao gênero do participante.



Fonte: elaboração própria.

O efeito da convencionalidade, relação entre regularidade, situação de uso e população (GIORA, 2003, p. 17) se verifica nos resultados de um modelo linear misto para estimar a nota de julgamento com a frequência com gênero feminino e frequência com gênero masculino, comando e sexo do participante] (fórmula = nota de julgamento \sim (frequência_F + frequência_M) * comando + sexo). O modelo incluiu participante como efeito aleatório (fórmula = ~ 1 | participante). O poder explicativo total do modelo é substancial (R^2 condicional = 0,46) e a parte relacionada apenas aos efeitos fixos (R^2 marginal é de 0,14). O coeficiente de correlação intraclasse é de 0,37, sugerindo baixa concordância e ressaltando o forte efeito da variável participante.

A interação entre frequência, familiaridade e convencionalidade direciona os processos de prototipia e estereotipia à medida que são “efeitos na cognição dos falantes sem manifestam na relação entre a frequência de ocorrência de uma variável, seu grau de previsibilidade de prototipia em um contexto e a percepção do falante” (FREITAG, 2018, p. 03). A interação entre frequência, familiaridade e convencionalidade direciona os processos de prototipia e estereotipia à medida que são “efeitos na cognição dos falantes sem manifestam na relação entre a frequência de ocorrência de uma variável, seu grau de previsibilidade de prototipia em um contexto e a percepção do falante” (FREITAG, 2018, p. 03).

O valor de intercepto do modelo, correspondente à nota de julgamento = 0, comando de julgamento da feminilidade da profissão e frequência como feminino e frequência como masculino baixa é 4,34 (SE = 0,26, $p < 0,001$). Neste modelo, o efeito de frequência como feminino alta é positivo e pode ser considerado muito grande e significativo ($\beta = 1,70$, SE = 0,16, $p < 0,001$), assim como o efeito de frequência como feminino altíssima ($\beta = 2,58$, SE = 0,16, $p < 0,001$). O efeito da frequência como masculino alta é negativo e pode ser considerado médio e significativo ($\beta = -1,17$, SE = 0,26, $p < 0,001$).

Tabela 2 - Modelo linear de efeitos mistos.

<i>Variáveis preditoras</i>	<i>Notas de Julgamento</i>			
	β	SE	<i>t</i>	<i>p</i>
(Intercepto)	4,34	0,26	16,97	<0,001
frequência_F [média]	0,16	0,14	1,13	0,257
frequência_F [alta]	1,70	0,16	10,53	<0,001
frequência_F [altíssima]	2,58	0,16	15,77	<0,001
frequência_M [média]	-0,13	0,21	-0,62	0,535
frequência_M [alta]	-0,19	0,24	-0,77	0,439

frequência_M [altíssima]	-1,17	0,26	-4,54	<0,001
comando [masculinidade da profissão]	0,21	0,36	0,59	0,552
sexo [homem]	-0,08	0,25	-0,33	0,739
frequência_F [média] *comando [masculinidade da profissão]	-0,02	0,20	-0,11	0,915
frequência_F [alta] *comando [masculinidade da profissão]	-2,88	0,23	-12,60	<0,001
Frequência_F [altíssima] * comando [masculinidade da profissão]	-4,48	0,23	-19,38	<0,001
frequência_M [média] *comando [masculinidade da profissão]	0,02	0,30	0,07	0,948
frequência_M [alta] *comando [masculinidade da profissão]	0,08	0,34	0,24	0,810
frequência_M [altíssima] *comando [masculinidade da profissão]	2,25	0,36	6,17	<0,001
<i>Efeitos aleatórios</i>				
σ^2	1,97			
τ_{00} participante	1,16			
ICC	0,37			
R ² marginal/ R ² condicional	0,13/0,45			
AIC	10544,81			

Fonte: elaboração própria.

O efeito da frequência como feminino alta em interação com o comando de julgamento da masculinidade da profissão é negativo e pode ser considerado grande e significativo ($\beta = -2,88$, SE = 0,23, $p < 0,001$), assim como a interação da frequência como feminino altíssima ($\beta = -4,48$, SE = 0,23, $p < 0,001$).

Já o efeito da frequência como masculino altíssima em interação com o comando de julgamento da masculinidade da profissão é positivo e pode ser considerado muito grande e significativo ($\beta = 2,25$, SE = 0,36, $p < 0,001$).

Os resultados apontam que as profissões altamente frequentes geram diferenças maiores no julgamento das profissões, independentemente de o comando ser feminino ou masculino. O efeito parece ser da frequência, e não do *priming*. A interação entre frequência, familiaridade e convencionalidade direciona os processos de prototipia e estereotipia (GIORA, 2003, p. 17). Esses “efeitos na cognição dos falantes se manifestam na relação entre a frequência de ocorrência de uma variável, seu grau de previsibilidade de prototipia em um contexto e a percepção do falante” (FREITAG, 2018, p. 03). Logo, os falantes marcam o gênero gramatical das profissões a partir das crenças e estereótipos

convencionalizados em sociedade a partir do quanto é mais frequente e familiar uma profissão para atender as características recorrentes para determinado estereótipo compartilhado entre os falantes, quanto mais frequente e familiar é vista e processada cognitivamente, mais forte é a representação social, pois assim, é mais fácil reconhecer o estereótipo (RÁCZ, 2013, p. 05).

O julgamento dos participantes para o gênero das profissões é baseado nos papéis estereotipados que já são construídos por um consciente coletivo na sociedade, como “eletricista” e “soldado”, para homens; e “babá” e “estilista”, para as mulheres. As profissões estão associadas ao efeito de saliência sociocognitiva: a exposição a um padrão de marcação de gênero desencadeiam fatores de frequência, familiaridade e convencionalidade, e ativam os estereótipos de gênero associados às funções e tarefas desempenhadas por cada falante executando em determinada profissão, em função do conhecimento das proporções de homens e mulheres que ocupam tais profissões (GYGAX, GARNHAM, DOEHREN, 2016).

Conclusões

A interação entre o julgamento das profissões comuns de dois gêneros no português e a frequência das palavras medidas pela marcação do gênero por meio do paradigma *Google as a corpus* (ROBB, 2003; DIEMER, 2008; SHA, 2010) pode ser uma evidência de que existe efeito entre a correspondência dos estereótipos com as informações de gênero gramatical. Os participantes da pesquisa (estudantes universitários) marcam a informação gramatical implícita de gênero nas profissões a partir das representações mentais de gênero nos estereótipos. Por exemplo, a profissão de “babá” está associada ao alto grau de feminilidade da profissão, enquanto “motorista”, “chofer” e “taxista” estão associadas ao alto grau de masculinidade das profissões. A marcação de gênero explicita a informação gramatical a partir da representação social dos estereótipos construídos de feminilidade e masculinidade na sociedade ocidental de tradição judaico-cristã.

O julgamento da masculinidade ou feminilidade de uma profissão depende do *priming* e da frequência dos nomes das profissões, o que pode ser considerado efeito de saliência sociocognitiva e frequência, associado às informações de estereotipia compartilhadas entre os grupos sociais.

REFERÊNCIAS

ARSENIEV-KOEHLER, Alina. Gender Equal or gender neutral? Gender connotations in gender fair language. In: *Proceedings* [...]. Conference on Computational Social Science IC2S2. University of Amsterdam: The Netherlands, p. 17-20, 2019.

AUGOUSTINOS, Martha; WALKER, Iain. *Social Cognition: a integrated introduction*. London: Sage publications, 1995.

FREITAG, Raquel Meister Ko. Saliência estrutural, distribucional e sociocognitiva. *Acta Scientiarum. Language and Culture*, v. 40, n. 2, p. 41173, 2018. Disponível em: <https://doi.org/10.4025/actascilangcult.v40i2.41173>. Acesso em: 05 de jan. 2020.

FREITAG, Raquel Meister Ko. Effects of the linguistics processing: palatals in Brazilian Portuguese and the sociolinguistic monitor. *University of Pennsylvania Working Papers in Linguistics*, v. 25, n 2, artigo 4, 2020. Disponível em: <https://repository.upenn.edu/pwpl/vol25/iss2/4>. Acesso em: 05 de jan. 2020.

BARR, Dale.; KEYSAR, Boaz. Making sense of how we make sense: The paradox of egocentrism in language use. In: COLSTON, Herbert L.; KATZ, Albert N. (eds). *Figurative language comprehension: Social and cultural influences*, London: Routledge, 2005.

BAAYEN, Harald. Probabilistic approaches to morphology. In: BOD, Rens; HAY, Jennifer; JANNEDY, Stefanie (eds.). *Probabilistic linguistics*. Mit Press, 2003, p.229-288.

BECHARA, Evanildo. *Moderna Gramática Portuguesa*. 37. e. d. Rio de Janeiro: Nova Fronteira, 2009.

BLUMENTHAL-DRAMÉ, Alice; HANULÍKOVÁ, Adriana; KORTMANN, Bernd. Perceptual linguistic salience: Modeling causes and consequences. *Frontiers in psychology*, v. 8, n. 411, 2017. Disponível em: <https://doi.org/10.3389/fpsyg.2017.00411>. Acesso em: 05 de jan. 2020.

BOCK, J. Kathryn; WARREN, Richard. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, v. 21, n. 1, p. 47-67, 1985. Disponível em: <http://www.linguisticsnetwork.com/wp-content/uploads/Bock-Warren-1986.pdf>. Acesso em: 05 de jan. 2020.

BOD, Rens; HAY, Jennifer; JANNEDY, Stefanie. *Probabilistic linguistics*. Mit Press, 2003.

BUTLER, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge, 1990.

BYBEE, Joan. From usage to grammar: The mind's response to repetition. *Language*, v. 82, n. 4, p. 711-733, 2006.

BYBEE, Joan; HOPPER, Paul. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins Publishing Company, 2001.

CAMARA Jr., Joaquim Mattoso. *Estrutura da língua portuguesa*. Petrópolis: Vozes, 1970.

CAMINO, Leôncio; TORRES, Ana Raquel Rosas. Origens e desenvolvimento da Psicologia Social. In: CAMINO, Leôncio; TORRES, Ana Raquel Rosas; PEREIRA, M. E. *Psicologia Social: temas e teorias*, Brasília: Technopolitik, p. 23-99, 2011.

CHIARCOS, Christian; CLAUS, Berry; GRABSKI, Michael. Saliency: Multidisciplinary perspectives on its function in discourse. Berlin: Walter de Gruyter, 2011.

CORBETT, Greville. Grammatical gender. In: BROWN, Keith. (org). *Encyclopedia of Language and linguistics*. 4 ed. Oxford: Elsevier, p. 749-756, 2007.

DIEMER, Stefan. Corpus linguistics with Google. *Proceedings [...]. ISLE 2*, USA: Boston, 2008. Disponível em: <http://www.bu.edu/isle/files/2012/01/Stefan-Diemer-Corpus-Linguistics-with-Google.pdf>. Acesso em: 08 de mai de 2020.

DIESSEL, Holger; HILPERT, Martin. Frequency effects in grammar. In: HARLEY, Heidi; MIYAGAWA, Shigeru (eds.). *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press, 2016. Disponível em <http://dx.doi.org/10.1093/acrefore/9780199384655.013.120>. Acesso em: 02 de abr. 2020.

FISK, Arthur.; SCHNEIDER, Walter. Memory as a function of attention, level of processing, and automatization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 10, n. 2, p. 181, 1984.

GARNHAM, Alan; OAKHILL, Jane; REYNOLDS, David. J. Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*. vol. 30, p. 439-446, 2002. Disponível em: <http://dx.doi.org/10.3758/BF03194944>. Acesso em: 05 de jan. 2020

GARNHAM, Alan.; DOEHREN, Sam; GYGAX, Pascal. True gender ratios and stereotype rating norms. *Frontiers in Psychology*, vol. 6, p. 01-07, 2015. Disponível em: <http://dx.doi.org/10.3389/fpsyg.2015.01023>. Acesso em: 05 de jan. 2020.

GEERAERTS, Dirk. Saliency phenomena in the lexicon: A typology. In ALBERTAZZI, Liliana (ed.). *Meaning and cognition*. Amsterdam/Philadelphia: John Benjamins, 2000, p. 79-101.

GIBSON, Edward; SCHÜTZE, Carson.; SALOMON, Ariel. The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, v. 25, n. 1, p. 59-92, 1996.

GIORA, Rachel. *On our mind: Saliency, context, and figurative language*. Oxford: Oxford University Press, 2003.

GIRAUDO, Helena; DAL MASO, Serena. Saliency of complex words and their parts: which comes first? In: BLUMENTHAL-DRAMÉ, Alice; HANULÍKOVÁ, Adriana; KORTMANN,

Bernd. *Perceptual Linguistic Saliency: Modeling Causes and Consequences*. Lausanne: Frontiers Media, 2017. Disponível em: <http://dx.doi.org/10.3389/978-2-88945-177-7>. Acesso em: 05 de jan. 2020

GYGAX, Pascal.; GARNHAM, Alan.; DOEHREN, Sam. What Do True Gender Ratios and Stereotype Norms Really Tell Us? *Frontiers in Psychology*, vol. 7, p.01-06, 2016. Disponível em: <http://dx.doi.org/10.3389/fpsyg.2016.01036>. Acesso em: 05 de jan. 2020.

HESTONE, Miles; FINCHAM, Frank; FOSTER, Jonathan. *Psychogy*. Blackwell: Oxford, 2005.

JASZCZOLT, Kasia.; ALLAN, Keith. Saliency and defaults in utterance processing. Berlin: Walter de Gruyter, 2011.

KECSKES, Istvan. Saliency in language production. In: JASZCZOLT, Kasia.; ALLAN, Keith (eds.). *Saliency and defaults in utterance processing*, Berlin: Walter de Gruyter, 2011, p. 81-102.

KECSKES, Istvan. The effect of saliency on shaping speaker's utterance. *Reti, saperi, linguaggi*, v. 4, n. 1, p. 5-34, 2017. Disponível em: <https://doi.org/10.12832/87352> . Acesso em: 05 de jan. 2020.

KENNISON, Sheila.; TROFE, Jessie. Comprehending pronouns: a role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, vol. 32, p. 355–378, 2003. Disponível em: <http://dx.doi.org/10.1023/A:1023599719948>. Acesso em: 05 de jan. 2020.

OAKHILL, Jane.; GARNHAM, Alan; REYNOLDS; David. Immediate activation of stereotypical gender information in reading. *Memory & Cognition*, vol. 33, p. 972-983, 2005. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/16496719>. Acesso em: 05 de jan.

PEIRCE, Jonathan. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, vol. 162, p.8-13, 2007. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2018741/>. Acesso em: 09 de jan. 2020.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. Disponível em: <http://www.Rproject.org>. Acesso em: 07 de jan. 2020.

RÁ CZ, Péter. *Saliency in sociolinguistics: a quantitative approach*. Berlin: Walter de Gruyter, 2013.

ROBB, Thomas. Google as a corpus tool. *ETJ Journal*, vol. 4, n. 1, p. 20-21, 2003.

SCHMID, Hans-Jörg; GÜNTHER, Franziska. Toward a unified socio-cognitive framework for saliency in language. *Frontiers in Psychology*, v. 7, p. 1110, 2016. Disponível em: <https://doi.org/10.3389/fpsyg.2016.01110>. Acesso em: 05 de jan. 2020.

SCHWINDT, Luiz Carlos. Exponência de gênero e classe temática em português brasileiro. *D.E.L.T.A.*, v. 34, n. 2, p.745-768, 2018. Disponível em: <http://dx.doi.org/10.1590/0102-44509154975775546>. Acesso em: 05 de jan. 2020.

SHA, Guoquan. Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. *Computer Assisted Language Learning*, v. 23, n. 5, p. 377-393, 2010. Disponível em: <https://doi.org/10.1080/09588221.2010.514576>. Acesso em: 05 de jan. 2020.

TAJFEL, Henri. Estereótipos sociais e grupos sociais. In: TAJFEL, Henri. *Grupos humanos e categorias sociais*. vol. 1. São Paulo: Livros Horizontes, 1981.

VILLALVA, Alina. *Estruturas Morfológicas: unidades e hierarquias nas palavras do português*. Rio de Janeiro: Fundação Calouste Gulbenkian, 2000.

**O PAPEL DA MORFOLOGIA CONSTRUCIONAL NA FORMAÇÃO DE
ANTROPÔNIMOS NEOLÓGICOS**
*THE ROLE OF CONSTRUCTIONAL MORPHOLOGY IN THE FORMATION OF
NEOLOGICAL ANTHROPONYMS*

Leticia Santos Rodrigues¹

RESUMO

Nesta pesquisa, intenta-se observar o papel da Morfologia Construcional na formação de antropônimos neológicos, segundo o aporte teórico da Linguística Cognitiva (BOOIJ, 2005; 2007; 2010; GOLDBERG, 1995; 2006). Os antropônimos em questão provêm do *corpus* analisado por Rodrigues (2019), a saber, as fichas de registros dos indivíduos que, ao fim do século XIX até início do século XXI, desejaram se vincular à Ordem Terceira do Carmo, localizada no Centro Histórico de Salvador/BA. O critério de seleção dos antropônimos é que estes apresentem, em suas construções, formativos vinculados parcial ou integralmente a uma origem germânica, em razão do fator histórico da ocupação da Península Ibérica pelos considerados “povos germânicos”, que influenciou o léxico antroponímico do Brasil não só pela presença desses formativos, mas também pela herança do modelo bitemático, a partir do qual a união de duas bases resulta em um novo item lexical. A consideração do critério neológico advém de os antropônimos não constarem em alguns dos principais dicionários onomásticos em língua portuguesa: Nascentes (1952), Guérios (1981) e Machado (2003). Dentre os resultados encontrados, é possível confirmar a presença do modelo bitemático nos antropônimos neológicos. Outro fato confirmado é o de que os indivíduos, quando expostos a um conjunto de antropônimos que se utilizam de dado formativo ou a um antropônimo muito produtivo, são capazes, por analogia, de depreender o modelo e reproduzi-lo ao criar novos antropônimos, como é possível demonstrar a partir de esquemas construcionais específicos, segundo os pressupostos da Linguística Cognitiva.

Palavras-chave: Onomástica. Morfologia Construcional. Linguística Cognitiva.

ABSTRACT

This research intends to observe the role of Construction Morphology in the formation of neological anthroponyms according to the theoretical framework of Cognitive Linguistics. The anthroponyms in question come from the corpus analyzed by Rodrigues (2019), namely, the records of individuals who, at the end of the 19th century until the beginning of the 21st century, wished to be linked to the *Ordem Terceira do Carmo* (Third Order of Carmel), located in the historic center of the city of Salvador (Bahia, Brazil). The anthroponyms selected present, in their constructions, formatives linked partially or wholly to a Germanic origin, due to the History of the Iberian Peninsula occupation by “Germanic people”, which influenced the anthroponomic lexicon of Brazil, not only by the presence of such formatives, but also by the inheritance of the bitematic model, from which the union of two bases results in a new lexical item. The neology criterion comes from the fact that anthroponyms are not included in some of the main onomastic dictionaries in Portuguese: Nascentes (1952), Guérios (1981) and Machado (2003). The results show it is possible to confirm the presence of the bithematic model in neological anthroponyms. Furthermore, when the individuals were exposed to a set of anthroponyms that use formative data or a very productive anthroponym, they are able to take the model and reproduce it by creating new anthroponyms by analogy. This is can be demonstrated from specific constructional schemes, according to the assumptions of Cognitive Linguistics.

Keywords: Onomastics. Construction Morphology. Cognitive Linguistics.

¹ Doutoranda no Programa de Pós-Graduação em Filologia e Língua Portuguesa da Universidade de São Paulo. E-mail: letisr@usp.br.

1 Introdução

A realização de um estudo linguístico pode perpassar certos domínios, a exemplo do sintático (organização dos itens lexicais numa frase e as relações que estes estabelecem entre si), fonético-fonológico (que se dedica, de modo geral, ao estudo dos sons), semântico (voltado para a questão do significado), pragmático (que se refere ao contexto de uso de uma dada palavra) e morfológico. Este último domínio, o morfológico, se volta para a análise da estrutura interna da palavra e é por esse viés que guiaremos nossas análises. Ademais, neste trabalho os itens lexicais a serem estudados se encontram no âmbito do léxico dos nomes próprios, nos direcionando, assim, para a Onomástica, considerada: “[...] área da linguística que se dedica ao estudo dos nomes próprios, quer se direcionem para as suas origens, quer para os seus processos de formação, quer para a sua organização no léxico das línguas e também no meio social” (RODRIGUES, 2016, p. 13).

A Onomástica se organiza em subáreas de acordo com a natureza de seus referentes, que podem ser: nomes de santos (Hagionímia), nomes de seres mitológicos (Mitonímia), nomes de plantas (Fitonímia), nomes de itens comerciais/marcas (Onionímia) etc. Contudo, as subáreas mais notadamente reconhecidas e investigadas são duas: a Toponímia – nomes de lugares – e a Antroponímia – nomes de pessoas –, segundo a qual nos orientaremos. Os objetos de estudo da Antroponímia são, assim, conhecidos como “antropônimos” e podem fazer menção aos nomes considerados oficiais (ou seja, registrados em cartório) – a saber, os prenomes e sobrenomes² que, unidos, compõem o ortônimo – ou aos alônimos, que são “[...] antropônimos que não correspondem com os nomes oficiais garantidos pela legislação e atribuídos ao indivíduo no registro civil” (AMARAL, 2011, p. 71-72), a exemplo dos apelidos, pseudônimos, heterônimos, nomes artísticos, dentre outros.

Para guiar nossas análises, nos centraremos no trabalho de Rodrigues (2019), no qual examina prenomes considerados neológicos que apresentem, em suas construções, formativos vinculados parcial ou integralmente a uma origem germânica. Salientamos que autores com formação não indo-europeístas³, como Nascentes (1952), Guérios (1981) e Machado (2003), muitas vezes rotulam seus étimos como “germânicos” pura e simplesmente, sem conjecturar que estes pudessem se tratar, por exemplo, de itens de “línguas germânicas antigas específicas” (como o gótico, antigo alto alemão, antigo saxão, antigo nórdico ou anglo-saxão). Dessa maneira, esses étimos considerados “germânicos” são, na verdade: a) pertencentes a uma língua germânica antiga específica não identificada; b) não

2 Na tradição portuguesa, os prenomes são todos os itens que antecedem os sobrenomes e que podem ser compostos por um, dois e, em casos mais raros, três elementos ou mais.

3 Antenor Nascentes direcionou seus estudos principalmente para a Filologia Românica, por isso a sua consideração como não indo-europeísta.

detectados, por sua difícil localização; c) erroneamente transcritos ou inexistentes. Logo, apesar de nos valermos das contribuições dos dicionários de Nascentes (1952), Guérios (1981) e Machado (2003) – entendendo que tais obras necessitam de revisões –, optamos, como em Rodrigues (2019), por usar parênteses uncinados em forma de aspas angulares «...» tanto na palavra «germânico» quanto nos supostos étimos quando estes forem incertos, uma vez que o uso do asterisco se refere a étimos reconstruídos por meio do método histórico-comparativo.

O motivo para a escolha de formativos vinculados parcial ou integralmente a uma origem germânica se deve ao modelo de formação desses nomes. Isso porque, segundo Piel (1989), no período indo-europeu, a nomeação entre os povos «germânicos» (assim como os gregos ou até antes), se utilizava, de modo geral, de uma formação bitemática, em que compostos personativos eram engendrados da união de dois itens do léxico comum, normalmente relacionados às características ou desejos dos pais para seus filhos, pois:

[...] a nomeação germânica durante parte da Idade Média era semanticamente motivada, com palavras relativas a ‘campos conceituais’ que se referiam a elementos próprios da cultura desses povos, como os de instrumentos de guerra, animais e sentimentos de conquista – coragem, fama e vitória –, além de adjetivos que denotavam qualidades físicas e morais (RODRIGUES, 2019, p. 76).

O objetivo deste trabalho será demonstrar que, segundo os princípios da Linguística Cognitiva, nossa base teórica, os indivíduos, ao depreenderem um padrão recorrente na língua, o reproduzem na criação de novos itens ou, no caso desta pesquisa, na criação de novos prenomes, seguindo o modelo de formação bitemática, coadunado com a hipótese proposta por Rodrigues (2019). Desse modo, passaremos ao entendimento do papel do Léxico e da Morfologia, da perspectiva cognitiva, na formação dos prenomes neológicos para, *a posteriori*, conhecer o *corpus* e seguir para as análises. *A posteriori*, na seção 3, serão apresentados o *corpus* e a metodologia adotada, de acordo com Rodrigues (2019). A análise dos dados se dará na seção 4, considerando os aspectos etimológicos, os possíveis prenomes que serviram como modelo e a proposição de esquemas construcionais que instanciem, por analogia, a formação de novos prenomes. Para concluir este trabalho, nossas considerações finais.

2 Breve conceito de Léxico e Morfologia sob o viés cognitivo

Antes de nos voltarmos para o domínio da Morfologia enquanto principal norteador desta pesquisa, entendemos ser necessário nos guiarmos pela concepção de léxico como um conjunto encontrado em toda e qualquer língua e que “[...] precisa ser interpretado como um componente indispensável da gramática, e não apenas como um rol de itens convencionalizados, fixos e armazenados na mente do falante, possibilitando, além da inovação criativa, uma organização hierarquizada” (RODRIGUES,

2019, p. 44). Outro ponto a ser observado no que se refere ao léxico é que:

A Linguística nunca cedeu ao léxico um lugar de centralidade, e desde as primeiras gramáticas, passando pelas históricas até as normativas tradicionais, o enfoque no léxico se resumia à análise de palavras primitivas e complexas já formadas e incorporadas ao léxico, e não havia interesse em observar a capacidade de se criarem novas palavras, nem de entender como essas palavras são armazenadas e se relacionam umas com as outras no cérebro dos falantes (SOLEDADE, 2018, p. 228).

Ao nos voltarmos especificamente para o processo de criação de novos antropônimos, e inscritos no paradigma da Linguística Cognitiva, tomamos como base a Morfologia Construcional, considerada “[...] um novo modelo teórico para o tratamento de palavras complexas que os falantes abstraem dos esquemas apreendidos com o uso da linguagem” (GONÇALVES, 2016, p. 9-10). Importante salientar que a expressão “Morfologia Construcional” foi proposta em 1987 por Corbin, contudo, sua abordagem estaria relacionada à perspectiva gerativista que “[...] reconhece uma competência derivacional, que permite a elaboração inconsciente de regras, e um saber lexical das convenções da língua, de aprendizagem idiossincrática” (GONÇALVES, 2006, p. 20). Já nos moldes da Linguística Cognitiva, o modelo da Morfologia Construcional foi idealizado e desenvolvido por Booij em três trabalhos (BOOIJ, 2005; 2007; 2010), nos quais são estabelecidas suas bases, com especial atenção para o papel do léxico e para o processo de formação de novas palavras. Na Morfologia Construcional de Booij também está presente a adaptação das ideias de Jackendoff (1997), com a Teoria da Entrada Plena (*Full Entry Theory*), e de Goldberg (1995), a respeito da Gramática de Construções, a partir da qual, ao contrário do pensamento gerativista, não haveria separação entre os componentes da gramática, em consonância com a Linguística Cognitiva. Logo:

Se a gramática é simbólica e, por isso, necessariamente evoca padrões instanciados pelas formas, suas motivações são intrinsecamente cognitivas e, em função de sua organização não modular, não há diferença de funcionamento entre os chamados níveis de descrição linguística (morfologia, léxico, sintaxe, fonologia) (GONÇALVES, 2016, p. 21).

Para resumir a questão, salientamos que a maior contribuição do aporte teórico da Linguística Cognitiva, em contraste com os princípios da teoria gerativa, está na consideração do aspecto semântico enquanto componente indispensável para o entendimento do léxico e de sua renovação, especificamente no caso deste trabalho, do ponto de vista onomástico. Isso porque, ao privilegiar a sintaxe como um módulo dissociado da fonologia e da semântica, a teoria gerativa não considera a interação entre a estrutura linguística como um todo e o uso conceptual, que são aspectos basilares na teoria cognitiva quando do julgamento do indivíduo como sujeito atuante no processo de renovação lexical a partir das interações que ele realiza com seu ambiente linguístico. Nesse sentido, Ferrari (2014, p. 13-14) confirma que:

[...] a teoria gerativa postula que o módulo cognitivo da linguagem é independente de outros módulos cognitivos (como o raciocínio matemático, a percepção etc.); além disso, no domínio da linguagem, reivindica-se a primazia do módulo sintático, que apresenta princípios próprios e independentes daqueles atuantes nos módulos fonológico e semântico, por exemplo. A Linguística Cognitiva, por sua vez, adota uma perspectiva não modular, que prevê a atuação de princípios cognitivos gerais compartilhados pela linguagem e outras capacidades cognitivas, bem como a interação entre os módulos da linguagem, mais especificamente, entre estrutura linguística e conteúdo conceptual.

Ao se pautar no uso, o processo de formação de novas palavras – e também a própria análise linguística desse processo – passa pela abstração, por parte do falante, de um padrão recorrente em vigor na língua, sobretudo por meio do mecanismo cognitivo da analogia que, para Bybee (2016, p. 27), “[...] é o processo pelo qual enunciados novos são criados com base em enunciados de experiências prévias”. Da perspectiva onomástica, vemos o processamento analógico como um recurso da cognição humana para a apreensão e, por conseguinte, criação de novos prenomes a partir de prenomes tradicionais, entendidos como os modelos. Contudo, para alcançarem essa condição de modelo, esses prenomes precisam apresentar frequência na língua, pois:

Cada uso de uma palavra ou construção aumenta a força de seu feixe de exemplares, tornando a palavra ou sintagma lexicalmente mais acessível. Em outros termos, a frequência de uso aumenta o nível de ativação latente de uma instância armazenada da construção. A força lexical maior de tal instância a torna mais propensa a ser acessada [...] (BYBEE, 2016, p. 125).

Como resultado de tais generalizações, o conjunto de novos antropônimos criados com dados formativos também pode ser expresso por esquemas. Assim, segundo Booij (2007, p. 34, tradução nossa⁴): “Os padrões de formação de palavras podem ser vistos como esquemas abstratos que generalizam conjuntos de palavras complexas existentes com uma correlação sistemática entre forma e significado. Esses esquemas também especificam como novas palavras complexas podem ser criadas”, como veremos nas seções dedicadas à apresentação e à análise do *corpus*. E complementa, em trabalho posterior, que a Morfologia Construcional: “[...] pressupõe que palavras complexas, isto é, as saídas de operações morfológicas, possam ser listadas no léxico. Os esquemas morfológicos, portanto, têm duas funções: eles expressam propriedades previsíveis de palavras complexas existentes e indicam como as novas podem ser cunhadas” (BOOIJ, 2010, p. 3, tradução nossa⁵). Ao analisarmos

4 No original: “Word formation patterns can be seen as abstract schemas that generalize over sets of existing complex words with a systematic correlation between form and meaning. These schemas also specify how new complex words can be created”.

5 No original: “[...] assumes that complex words, i.e. the outputs of morphological operations, can be listed in the lexicon. Morphological schemas therefore have two functions: they express predictable properties of existing complex words and indicate how new ones can be coined”.

os prenomes recolhidos do *corpus* e também o comportamento de cada formativo em prenomes considerados tradicionais⁶, será possível propor possíveis esquemas para os formativos em questão, demonstrando como esse processo de abstração e criação ocorre também com itens onomásticos⁷.

Por fim, ao invés de “morfema” – conceito tradicionalmente empregado no âmbito dos trabalhos em Morfologia –, adotaremos o termo “formativo” para a unidade mínima, fruto da nossa segmentação mórfica, seguindo Soledade e Lopes (2015), Gonçalves (2016) e Simões Neto e Soledade (2018). Pois, além de não desejarmos nos comprometer com paradigmas teóricos já defasados – que entendem o morfema como a unidade mínima de significado e recorrente, desconsiderando suas particularidades – e dissonantes do escopo teórico da Linguística Cognitiva, acreditamos que o estudo onomástico apresenta particularidades em relação ao estudo do léxico comum. Para citar algumas dessas especificidades, além do reconhecido do uso obrigatório de maiúsculas para os nomes próprios (quanto ao aspecto gráfico), mencionamos as contribuições de Ullmann (1967), que estabelece cinco critérios de diferenciação entre nomes próprios e comuns, a saber: a) identificação: corresponde à ideia de que um nome próprio identifica o indivíduo, como uma espécie de “etiqueta”; b) som distintivo: afirma que sons referentes aos nomes próprios teriam, em si, maior capacidade de prender a atenção dos utentes do que nomes comuns, que são mais gerais; c) critérios gramaticais: são critérios variáveis entre idiomas, regiões e épocas, mas um exemplo é a não pluralização de nomes próprios, exceto quando estes passam a figurar como nomes comuns; d) unicidade: é a ideia de que o nome próprio torna seus indivíduos únicos em relação aos demais. Contudo, a ideia de unicidade é problemática ao considerarmos o grande número populacional e, conseqüentemente, os inúmeros casos de homonímia; e e) designação contra conotação: para Ullmann (1967), por serem designativos, os nomes próprios seriam desprovidos de significado, diferentemente dos nomes comuns – o que discordamos, pois nesse caso o autor parece se referir apenas ao valor semântico dos nomes próprios em sua perspectiva sincrônica, descartando o valor etimológico e documental.

6 Os critérios de classificação dos prenomes como tradicionais ou neológicos foram estabelecidos nesta pesquisa – assim como em Rodrigues (2016; 2019) – de acordo com o projeto Todos os Nomes, desenvolvido na Universidade Federal da Bahia e coordenado pelas professoras doutoras Ariadne Almeida, Juliana Soledade e Tânia Lobo, entre os anos de 2007 e 2009, no âmbito do Programa para a História da Língua Portuguesa (PROHPOR). Para confirmar o *status* neológico de um prenome é preciso que ele esteja ausente nos dicionários onomásticos de Nascentes (1952), Machado (2003) – com o acréscimo do dicionário de Guérios (1981).

7 Exemplos de trabalhos que vêm se pautando na Linguística Cognitiva e no uso de esquemas construcionais para explicar os processos de formação de palavras estão em Gonçalves (2016), Gonçalves e Almeida (2014), Rodrigues (2016; 2019), Simões Neto (2016), Simões Neto e Soledade (2018) e Soledade (2018), para citar alguns.

Nos aprofundando mais no critério de designação contra conotação, proposto por Ullmann (1967), entendemos que, mesmo apresentando essas particularidades semânticas quando comparados com os itens do léxico comum – tanto os nomes concretos quanto os abstratos –, não é seguro concordar que os antropônimos sejam itens esvaziados de sentido. O que propomos, portanto, é que os itens pertencentes ao léxico onomástico, de modo geral, quando não transparentes quanto à sua carga semântica do ponto de vista sincrônico, caminham para um processo de opacidade semântica, ou seja, em que não há conhecimento evidente da carga significativa desses nomes por parte dos indivíduos, visto que esse conhecimento poderia apenas ser recuperado a partir da pesquisa/consulta etimológica. Assim, dificilmente saberíamos que prenomes como “Bruno” ou “Letícia” remontam, respectivamente, ao significado de “moreno” (MACHADO, 2003, p. 289) ou “alegria” (GUÉRIOS, 1981, p. 161), pois são informações advindas de obras etimológicas e/ou onomásticas. Nesse sentido, nos diz Carvalhinhos e Antunes (2007, p. 118):

A opacidade ocorre por dois motivos principais: o primeiro, mais óbvio, é por desconhecimento da língua em questão, o que leva à não decodificação ou não complementação do percurso [de] decodificação daquele nome. No segundo motivo, o tempo e a sobreposição de camadas linguísticas concorrem para a opacidade do nome [...].

Além do mais, em Linguística Cognitiva, todo item lexical é dotado de sentido e, “[...] por muitas vezes aludirem a sincronias pretéritas, nem sempre passíveis de decodificação por seu utente hodierno, os antropônimos guardam, em si, informações de significado acessadas somente por meio da pesquisa etimológica e documental” (RODRIGUES, 2019, p. 24). Encerrando o assunto, por ora, elencamos alguns dos principais problemas que surgem a partir da total desconsideração do conteúdo semântico dos antropônimos: a) a renúncia ao significado etimológico, conhecido por meio das pesquisas históricas e presente nas obras onomásticas; b) a negação de aspectos semânticos, como a referência ao gênero masculino ou feminino; c) a associação óbvia, da perspectiva sincrônica, de certos antropônimos com nomes comuns, como nos casos de “Rosa”, “Hortênsia”, “Sol” etc., além da associação, em alguns casos, de antropônimos com nomes considerados tabus (por motivos diversos), como “Marciano”, “Frígida”, “Bucetíldes”, o que, se não fosse por sua carga semântica, não seriam considerados constrangedores ou, até mesmo, jocosos.

Ao propor esquemas construcionais para os formativos analisados, entenderemos como funciona essa questão do significado para os prenomes, visto que esquemas são generalizações que apresentam pareamento entre forma e significado. Para tanto, seguiremos com a apresentação do *corpus* e a metodologia utilizada.

3 Apresentação do *corpus* e dos métodos

O *corpus* selecionado nesta pesquisa é um recorte do que foi utilizado por Rodrigues (2019) em sua pesquisa de mestrado e se refere aos arquivos da Ordem Terceira do Carmo⁸, localizada no Centro Histórico da cidade de Salvador, Bahia. Essa documentação é uma fonte histórica que registra as ações dos indivíduos que, no passado, desejaram se vincular à Ordem. A escolha do *corpus* se deu por alguns motivos, dentre eles: a) o local, pois segundo a hipótese formulada por Rodrigues (2019), há incremento neológico no quadro onomástico brasileiro após a instauração do Decreto nº 9.886 e a abolição da escravatura, ambos em 1888. Isso porque haveria a vontade de parte da população negra e descendente de africanos de buscar uma construção identitária particular, uma vez que estes não se viam representados pela herança onomástica recebida dos seus senhores, partindo para um caminho alternativo ao criar, por meio de mecanismos cognitivos e de suas experiências lexicais, novos nomes. As consequências de tais acontecimentos poderiam ser mais facilmente observadas no estado da Bahia, onde havia alto índice da população negra (RODRIGUES, 2019); b) o período histórico, por remontar ao ano de 1893 (ano do primeiro registro) até 2014 (ano do último registro), ou seja, poucos anos após a instauração do Decreto nº 9.886 e da abolição da escravatura até o período atual; c) pela força histórica e documental que guardam os documentos vinculados a instituições religiosas, muitas vezes relegados em estudos linguísticos e filológicos. Infelizmente, devido à ausência de informações, não foi possível traçar um perfil sócio-histórico desses indivíduos, o que não caracterizou nenhum demérito à pesquisa.

Esses arquivos são compostos por fichas de registro dos indivíduos que, no passado, desejaram se vincular à Ordem e constituem os 10 tomos do Livro dos Termos dos Irmãos, tendo sido analisados por Rodrigues (2019) os Livros de número 7, 8, 9 e 10, percorrendo fins do século XIX até o início do século XXI. Cada livro possui, respectivamente, 796, 500, 500 e 164 fichas, totalizando 1960 antropônimos.

Os critérios para caracterização do prenome com o *status* de neológico foram os de não constarem nos principais dicionários etimológicos portugueses, a saber: Nascentes (1952), Guérios (1981) e Machado (2002). Após, foram selecionados todos os prenomes neológicos que apresentassem em sua construção um ou mais formativos que se vinculassem parcial ou integralmente a uma origem germânica. Depois de detectar os formativos em questão, os prenomes passaram por uma segmentação mórfica, a fim de identificar possíveis usos relacionados à herança bitemática e à posição ocupada pelos formativos, como realizado em Rodrigues (2019).

8 As Ordens chamadas “terceiras” são associações de leigos católicos em devoção. Um exemplo é a Ordem Terceira do Carmo, também conhecida como “Ordem dos Carmelitas”.

Neste trabalho, analisamos apenas dois dos formativos «germânicos» de maior frequência no *corpus* utilizado por Rodrigues (2019), a saber: ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD e ILD- ~ -ILD⁹, totalizando 17 prenomes (incluindo os prenomes repetidos, por razões quantitativas), encontrados nos Livros 9 e 10, conforme demonstrado no Quadro 1.

Quadro 1 – Prenomes neológicos segundo os formativos mais recorrentes que se vinculam parcial ou integralmente a uma origem germânica

Livro	Prenome	Ano de registro
10	Avanildes	1985
9	Crisaldete	1957
9	Eurivalda	1957
9	Everaldina	1955
9	Hildélia ~ Hyldélia ¹⁰	1964
9	Hildete	1957
9	Hildete	1978
8	Hildeth ~ Hyldeth	1926
9	Hyldéa	1976
9	Ironilde	1977
9	Marinaldo	1975
9	Rahilda	1961
10	Railda	1987
10	Valdiva	1984
9	Waldelir	1942
10	Waldete	1980
9	Wiveraldo	1977

Fonte: Adaptado de Rodrigues (2019).

Portanto, a seguir serão analisados oito prenomes correspondentes aos formativos ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD e nove prenomes correspondentes ao formativo ILD- ~ -ILD. O foco das análises se centrará: a) na posição que estes ocupam nas construções; b) nos possíveis usos do modelo bitemático e; c) na proposição dos respectivos esquemas.

4 Análise do *corpus*: propostas de segmentação mórfica e de esquemas construcionais

Além das propostas de segmentação mórfica referentes aos 17 prenomes selecionados,

9 Utilizamos o elemento gráfico “~” para indicar que o mesmo formativo pode aparecer ocupando posições diferentes no composto neológico, como ficará mais claro após a segmentação mórfica e a proposição dos esquemas construcionais.

10 Novamente nos utilizamos do elemento gráfico “~”, dessa vez para indicar incongruências verificadas nas fichas de registro que compõem o *corpus*. Isso porque, por se tratar de um documento manuscrito e, em muitos casos, escrito a duas mãos, apresenta divergências na grafia dos antropônimos (de acordo com o primeiro campo, o de identificação – geralmente preenchido por um membro da Ordem – e o segundo campo, o de assinatura – preenchido pelo indivíduo). Optamos, então, por apresentar as duas grafias encontradas, entendendo que tal fato não compromete as análises.

apresentaremos também os dois formativos de acordo com os seus possíveis significados etimológicos e com os nomes tradicionais em que aparecem, já que acreditamos, como dito, que os falantes, após serem expostos a um dado conjunto de antropônimos (ou a um que seja muito frequente no seu quadro onomástico), depreendem o modelo e, assim, criam novos antropônimos.

4.1 OS FORMATIVOS ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD

De acordo com Piel (1933), a partir do prenome “Rabaldo”, os formativos ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD seriam derivados do verbo «*waldan*», “governar” e, por isso, apareceram agrupados nas nossas análises. Esse étimo foi apresentado por Nascentes (1952) como oriundo do gótico, mas como explicado anteriormente, serão usadas aspas angulares sempre que nos utilizarmos de étimos não comprovados – o que não invalida a análise. Dentre os prenomes encontrados em Nascentes (1952) com esses formativos, podemos citar: a) ocupando a posição inicial (ou com base mais à esquerda): “Aldobrando”, “Aldonça” e “Valdomiro”; b) ocupando a posição final (ou com base mais à direita): “Adroaldo”, “Aguinaldo”, “Arnaldo”, “Beraldo”, “Bernaldo”, “Clodoaldo”, “Euvaldo”, “Evaldo”, “Geraldo”¹¹, “Giraldo”, “Heraldo”, “Osvaldo”, “Reginaldo”, “Reinaldo” e “Vivaldo”; c) ocupando posição absoluta: “Aldo” e “Valdo”.

Ao consultar Nascentes (1952), portanto, já é possível perceber que ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD apresentam mobilidade no composto, aparecendo tanto em posição inicial quanto final e absoluta. Assim também ocorreu com os oito prenomes do *corpus*, como observamos no Quadro 2.

Quadro 2 – Segmentação mórfica para os prenomes considerados neológicos com os formativos ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD

Prenome	Proposta de segmentação mórfica
Crisaldete	[CRIS- + -ALD(ETE)]
Eurivalda	[EURI- + -VALD(A)]
Everaldina	[EVER- + -ALD(INA)]
Marinaldo	[MARI- + -NALD(O)]
Valdiva	[VALD- + -IVA] ¹²
Waldelir	[WALD- + -ELIR]
Waldete	[WALD- + -ETE]
Wiveraldo	[WIVER- + -ALD(O)]

Fonte: Adaptado de Rodrigues (2019).

11 Piel (1933) propõe que o formativo -ALD no prenome “Geraldo” não é o mesmo que apresentamos referente a «*waldan*», mas sim referente a «*hartus*», significando “duro”, além de afirmar que não há outros nomes visigóticos com tal componente. Porém, devido à imprecisão do étimo e da ampla difusão que o prenome “Geraldo” possui no quadro antroponímico brasileiro, sendo um expressivo representante do modelo que envolve esse formativo, o mantivemos nesta pesquisa, também considerando os prenomes “Geralda” e “Geraldino” (RODRIGUES, 2019).

12 Sem desconsiderar a possibilidade de que tal prenome seja formado pela união das bases livres [VAL + DIVA], o que só o contato com o responsável pela criação antroponímica poderia confirmar – o que não é possível.

Destarte, temos três prenomes em posição inicial ou com base mais à esquerda (“Valdiva”, “Waldelir” e “Waldete”) e cinco em posição final ou com base mais à direita (“Crisaldete”, “Eurivalda”, “Everaldina”, “Marinaldo” e “Wiveraldo”), como demonstra a segmentação mórfica. Outro aspecto que se pode depreender é a presença do modelo bitemático em quase todos os prenomes do *corpus* (exceto “Waldete”, em que o formativo VALD- aparece unido a um diminutivo¹³). A afirmação de que se trata de duas bases vem da depreensão feita prenome a prenome, que pode ser realizada por meio da técnica de comutação quando, por exemplo, ao elegermos o prenome “Crisaldete”, encontramos também os prenomes “Crisália”, “Crisalva”, “Crisana”, “Criselda”¹⁴ etc.

Desse modo, após conhecermos os prenomes tradicionais recolhidos em Nascentes (1952) e também os prenomes neológicos recuperados no *corpus*, propomos os seguintes esquemas construcionais de acordo com a(s) posição(ões) ocupada(s) pelos formativos:

I. Posição inicial ou base mais à esquerda:

[Ald-X]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

[Nald-X]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

[Vald-X]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

II. Posição final ou base mais à direita:

[X-ald]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

[X-nald]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

[X-vald]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

III. Posição absoluta:

[Ald(a/o)]_{NP} ↔ [nome de pessoa]_{NP}

[Nald(a/o)]_{NP} ↔ [nome de pessoa]_{NP}

[Vald(a/o)]_{NP} ↔ [nome de pessoa]_{NP}

O símbolo de seta dupla (↔) remete à indexação de natureza categorial que indica a formação de um nome próprio (NP) e a contraparte semântica (também NP), estabelecendo, portanto, a relação entre forma e significado. No caso dos formativos ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD apresentados em posição absoluta, é possível conhecer, além da informação de que se trata de um nome próprio,

13 Nesse caso, o diminutivo aparece com um significado considerado opaco.

14 Tais prenomes podem ser confirmados na página Nomes no Brasil, do Censo 2010 do Instituto Brasileiro de Geografia e Estatística. Disponível em: <https://censo2010.ibge.gov.br/nomes/#/search>.

se se refere à pessoa do gênero¹⁵ feminino ou masculino, pois em português frequentemente ocorre essa demarcação com as vogais “a” e “o”, o que não ocorre quando esses formativos aparecem em posição inicial.

4.2 O FORMATIVO ILD- ~ -ILD

Encontramos a informação de que o formativo ILD- ~ -ILD está vinculado a uma origem «germânica», do antigo alto alemão **hilti*¹⁶ (HAHN, [200-]). Quanto aos prenomes dicionarizados com esse formativo, verificamos em Nascentes (1952): a) em posição inicial (ou com base mais à esquerda): “Hildeberto”, “Hildebrando”, “Hildegarda”, “Hildemar”, “Hilderico”, “Ildefonso”; b) em posição final (ou com base mais à direita): “Batilde”, “Brunilda”, “Cassilda”, “Clotilde”, “Cremilda”, “Heremildo”, “Giselda”, “Leonilda”, “Matilde”, “Ragnilda”, “Tusnelda”; c) ocupando posição absoluta: “Hilda” e “Hildo”. Para verificar a mobilidade do formativo nos prenomes encontrados no *corpus*, passaremos à apresentação do Quadro 3.

Quadro 3 – Segmentação mórfica para os prenomes considerados neológicos com o formativo ILD- ~ -ILD

Prenome	Proposta de segmentação mórfica
Avanildes	[AVA(N)- + -ILDES]
Hildete	[HILD- + -ETE]
Hildeth ~ Hyldeth	[HILD- + -ETH ~ HYLD- + -ETH]
Hildélia ~ Hyldélia	[HILD- + -ÉLIA ~ HYLD- + -ÉLIA]
Hyldéa	[HYLD- + -ÉA]
Ironilde	[IRO(N)- + -ILDE]
Rahilda	[RA(H)- + -ILDA]
Railda	[RA- + -ILDA]

Fonte: Adaptado de Rodrigues (2019).

Portanto, assim como os formativos ALD- ~ -ALD/NALD- ~ -NALD/VALD- ~ -VALD, vemos que ILD- ~ -ILD também apresenta mobilidade, tanto nos prenomes dicionarizados quanto nos prenomes neológicos, ocupando a posição inicial ou com base mais à esquerda (“Hildete”, “Hildeth” ~ “Hyldeth”, “Hildélia” ~ “Hyldélia” e “Hyldéa”) e a posição final ou com base mais à direita (“Avanildes”, “Ironilde”, “Rahilda” e “Railda”). Essa mobilidade ratifica a autonomia do utente ao criar prenomes, entendendo que este, ao depreender o funcionamento de dado formativo antropônimo na língua, pode reinterpretá-lo (ou não) quanto à posição que ocupará no novo prenome, de acordo com os

15 Salientando que o uso da palavra “gênero” se refere ao gênero gramatical, sem desprezar questões mais complexas de cunho sociopsicológico que dizem respeito aos nomes de pessoas e ao próprio termo em questão.

16 A diferença no uso dos sinais gráficos em **hilti* e «*waldan*» tem um motivo. No caso de «*waldan*», esse étimo foi retirado de Nascentes (1952) que, conforme explicado, não é um linguista indo-europeísta, portanto, suas proposições de étimos não se pautam em reconstruções advindas do método histórico-comparativo. O uso do asterisco indica que se trata de um étimo reconstruído por tal método e, por isso, foi usado em **hilti*.

princípios da Linguística Cognitiva. Em resumo, se o indivíduo conhece o dado formativo em posição inicial, ele pode optar por mantê-lo nessa mesma posição ou movê-lo para a posição final, formulando diferentes esquemas.

Confirmamos também a presença do modelo bitemático na estruturação desses prenomes, mesmo em casos menos salientes, como os dos prenomes “Avanildes” (“Avanice”, “Avanil”, “Avanir”) e “Ironilde” (“Ironalda”, “Ironice”, “Ironir”). Logo, após depreensão dos prenomes tradicionais e dos prenomes neológicos, propomos os seguintes esquemas construcionais para o formativo ILD- ~ -ILD:

I. Posição inicial ou base mais à esquerda:

[Ild-X]_{NP} ↔ [nome de pessoa associado ao formativo X]_{NP}

II. Posição absoluta:

[Ild(a/e/es/o)]_{NP} ↔ [nome de pessoa]_{NP}

Novamente ocorre a indexação da natureza categorial com a contraparte semântica, com a vogal final “a” e “o” determinando o gênero da maioria dos prenomes. Apresentada a análise do *corpus*, passaremos para as nossas considerações finais.

Considerações finais

Após as análises envolvendo alguns dos prenomes do *corpus* de Rodrigues (2019), foi possível confirmar a hipótese já levantada pela autora de que a criação de prenomes neológicos no quadro onomástico baiano¹⁷ se vale, em grande medida, do modelo bitemático, com a união de duas bases já conhecidas por meio dos prenomes tradicionais – como demonstramos por meio da segmentação mórfica. Ainda, em consonância com a Linguística Cognitiva, acreditamos que o falante, ao ser exposto a um amplo conjunto de antropônimos ou a um antropônimo que seja muito produtivo na língua, é capaz de depreender, por analogia, esquemas que permitam a criação de novos. Esses esquemas exprimem as generalizações entendidas a partir do uso, somado ao fato de que, ao considerar a contraparte semântica dos antropônimos, estes estabelecem uma relação não arbitrária entre forma e significado na construção.

Em suma, diferentemente de outras teorias, ressaltamos a importância da Linguística Cognitiva e, por conseguinte, da Morfologia Construcional que, ao se pautarem no uso e na cognição do

17 Apesar de supormos que, apresentando variabilidade em termos quantitativos, essa hipótese se aplique a todo o território nacional – o que apenas um *corpus* mais abrangente seria capaz de determinar.

indivíduo, valorizam seus conhecimentos experienciados, reconhecidos por meio da criação lexical, caracterizando os processos linguísticos como sendo “[...] cognitivos, sociointeracionais, culturais e históricos” (SOLEDADE, 2018, p. 225). Tal ponto de vista vai de encontro a argumentos ultrapassados que veem os nomes apenas como etiquetas identificatórias dos indivíduos, ignorando os saberes históricos, políticos e culturais intrínsecos a eles.

REFERÊNCIAS

AMARAL, Eduardo Tadeu Roque. Contribuições para uma tipologia de antropônimos do português brasileiro. *Alfa: revista de linguística*, São Paulo, v. 55, n. 2, p. 63-82, 2011.

BOOIJ, Geert. Compounding and derivation: evidence for construction morphology. In: DRESSLER, Wolfgang U. et al. (ed.). *Morphology and its demarcations*. Amsterdam: John Benjamins Publishing Company, 2005. p. 109-131.

BOOIJ, Geert. Construction morphology. *Language and Linguistics Compass*, United Kingdom, v. 4, n. 7, p. 543-555, 2010. Disponível em: <https://geertbooij.files.wordpress.com/2014/02/booi-2010-construction-morphology-lg-linguistics-compass.pdf>. Acesso em: 20 out. 2019.

BOOIJ, Geert. Construction morphology and the lexicon. In: MONTERMINI, Fabio; BOYÉ, Gilles; HATHOUT, Nabil (eds.). *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville: Cascadilla Press, 2007, p. 34-44.

BYBEE, Joan. *Língua, uso e cognição*. Tradução Maria Angélica Furtado da Cunha. São Paulo: Cortez, 2016.

CARVALHINHOS, Patrícia de Jesus; ANTUNES, Alessandra Martins. Princípios teóricos de toponímia e antroponímia: a questão do nome próprio. In: CONGRESSO NACIONAL DE LINGUISTICA E FILOGIA, 2007, Rio de Janeiro. *Livro dos minicursos [...]*. Rio de Janeiro: Cifefil, 2007. p. 108-121.

DICK, Maria Vicentina de Paula do Amaral. *Toponímia e antroponímia no Brasil*. 3. ed. São Paulo: Coletânea de Estudos, 1992.

FERRARI, Lilian. *Introdução à linguística cognitiva*. São Paulo: Contexto, 2014.

GOLDBERG, Adele. *Constructions: a Construction Grammar approach to argument structure*. Chicago/London: The University of Chicago Press, 1995.

GONÇALVES, Carlos Alexandre. *Morfologia construcional: uma introdução*. São Paulo: Contexto, 2016.

GONÇALVES, Carlos Alexandre; ALMEIDA, Maria Lúcia Leitão de. Morfologia construcional: principais ideias, aplicação ao português e extensões necessárias. *Alfa: revista de linguística*, São Paulo, v. 58, n. 1, p. 165-193, 2014. Disponível em: <http://seer.fclar.unesp.br/alfa/article/view/5771/4922>. Acesso em: 18 nov. 2017.

GUÉRIOS, Rosário Farâni Mansur. *Dicionário etimológico de nomes e sobrenomes*. 3. ed. São Paulo: Ave Maria Ltda, 1981.

HAHN, Simone. *Hildegart [S.l.]*: Althochdeutscher Namenbaukasten, [200-]. Disponível em: <https://homepages.uni-tuebingen.de/henrike.laehnemann/namen.html#h1>. Acesso em: 20 abr. 2020.

JACKENDOFF, Ray. *The architecture of the language faculty*. Cambridge: MIT Press, 1997.

MACHADO, José Pedro. *Dicionário onomástico etimológico da língua portuguesa*. Lisboa: Horizonte: Confluência, 2003. v. 2.

NASCENTES, Antenor. *Dicionário etimológico da língua portuguesa*. Rio de Janeiro: Francisco Alves, 1952.

PIEL, Joseph Maria. *Estudos de linguística histórica galego-portuguesa*. Lisboa: IN/CM, 1989.

PIEL, Joseph Maria. *Os nomes germânicos na toponímia portuguesa*. Boletim de Filologia. Lisboa: Centro de Estudos Filológicos, 1933.

RODRIGUES, Letícia Santos. *Neologia antroponímica: o que os nomes de origem germânica têm a nos dizer?*. 2019. Dissertação (Mestrado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019.

RODRIGUES, Letícia Santos. *Neologismos antroponímicos com base na utilização de formativos germânicos no Brasil*. 2016. Trabalho de Conclusão de Curso (Bacharelado em Letras) – Instituto de Letras, Universidade Federal da Bahia, Salvador, 2016.

RODRIGUES, Letícia Santos; SOLEDADE, Juliana. Germanismos e a contribuição para a antroponímia brasileira. *Revista Hyperion*, Salvador, n. 8, p. 75-90, 2016. Disponível em: <http://>

www.portalseer.ufba.br/index.php/revistahyperion/article/view/17018/11379. Acesso em: 18 out. 2019.

ROSA, Maria Carlota. *Introdução à morfologia*. São Paulo: Contexto, 2000.

SANDMANN, Antônio José. *Morfologia geral*. 3. ed. São Paulo: Contexto, 1997.

SANDMANN, Antônio José. *Morfologia lexical*. São Paulo: Contexto, 1992.

SIMÕES NETO, Natival Almeida. *Um enfoque construcional sobre as formações X-eir-: da origem latina ao português arcaico*. 2016. 2 v. Dissertação (Mestrado em Língua e Cultura) – Instituto de Letras, Universidade Federal da Bahia, Salvador, 2016.

SIMÕES NETO, Natival Almeida; SOLEDADE, Juliana. Nomes masculinos X-son na antroponímia brasileira: uma abordagem morfológica, histórica e construcional. *Revista de Estudos da Linguagem*, Belo Horizonte, v. 26, n. 3, p. 1295-1350, 2018. Disponível em: http://periodicos.letras.ufmg.br/index.php/relin/article/view/12678/pdf_1. Acesso em: 23 out. 2019.

SOLEDADE, Juliana. Por uma abordagem cognitiva da morfologia: revisando a morfologia construcional. In: ALMEIDA, Aurelina Ariadne Domingues; SANTOS, Elisângela Santana dos (org.). *Linguística Cognitiva: redes de conhecimento d'aquém e d'além-mar*. Salvador: EDUFBA, 2018, p. 225-258. Disponível em: https://repositorio.ufba.br/ri/bitstream/ri/28269/1/linguistica-cognitiva_miolo.pdf. Acesso em: 24 out. 2019.

SOLEDADE, Juliana; LOPES, Mailson. Uma proposta de revisão do conceito de morfema. In: ALMEIDA, Aurelina Ariadne Domingues; SANTOS, Elisângela Santana dos; SOLEDADE, Juliana (org.). *Saberes lexicais: mundos, mentes e usos*. Salvador: EdUFBA, 2015. p. 429-461.

ULLMANN, Stephen. *Semântica: uma introdução à ciência do significado*. Tradução Mateus, J. A. Osorio. 2. ed. Lisboa: Calouste-Gulbenkian, 1967.

VIARO, Mário Eduardo; FERREIRA, Michael; GUIMARÃES FILHO, Zwinglio. Derivação ou terminação: limites para a semântica, lexicologia e morfologia históricas. In: VIARO, Mário Eduardo (org.). *Morfologia histórica*. São Paulo: Cortez Editora, 2014, p. 58-105.

**ESTUDO SOCIOFONÉTICO DO PORTUGUÊS FALADO NA AMAZÔNIA: UMA
COMPARAÇÃO INTERDIALETAL ENTRE BELÉM E MACAPÁ**
*SOCIOPHONETIC STUDY OF PORTUGUESE SPOKEN IN THE AMAZON: AN
INTERDIALETAL COMPARISON BETWEEN BELÉM AND MACAPÁ*

Rosinele Lemos e Lemos¹
Camila Roberta dos Santos Brito²

RESUMO

Este artigo apresenta uma comparação interdialetoal entre as variedades de Belém (PA) (BRITO, 2014) e Macapá (AP) Lemos (em andamento), com o objetivo de investigar a variação geoprosódica do português falado na Amazônia Oriental. Foram analisados vocábulos oxítonos presentes em sentenças declarativas neutras e interrogativas totais, com o intuito de observar a variação entoacional da frequência fundamental (F0) em posição nuclear dos enunciados. A análise tem como base a Sociofonética (THOMAS, 2011), uma vez que compreende uma análise acústica da fala dentro do contexto social dos informantes. Trata-se de uma pesquisa do projeto AMPER-Amazônia, vinculada ao AMPER-POR. O *corpus* analisado totalizou 432 dados (3 sentenças x 2 modalidades x 2 variedades x 3 repetições x 12 informantes). O tratamento dos dados abarcou seis etapas: codificação das repetições; isolamento, segmentação automática no PRAAT; aplicação do *script* no PRAAT; seleção das três melhores repetições e aplicação da interface MatLab. Os resultados deste inédito e relevante estudo revelaram características prosódicas nos falares de belenenses e macapaenses ao apontarem contornos prosódicos similares da F0, na sílaba tônica final nas duas modalidades analisadas, porém evidenciaram-se traços prosódicos mais regulares nas interrogativas totais.

Palavras-chave: Análise prosódica. Sociofonética. Português Brasileiro. Projeto AMPER.

ABSTRACT

This paper presents an interdialectal comparison between Belém (PA) (BRITO, 2014) and Macapá (AP) Lemos (in progress) varieties, aiming at investigating the geoprosodic variation in the Portuguese spoken in Eastern Amazonia. Oxytone words in neutral declarative and total interrogative sentences were analyzed in order to observe the intonational variation of the fundamental frequency (F0) in the nuclear position of the utterances. The analysis is based on Sociophonetics (THOMAS, 2011), since it addresses an acoustic analysis of speech within the social context of the informants. This research is from the AMPER-Amazonia project, associated with the AMPER-POR project. The analyzed *corpus* totalized 432 observations (3 sentences x 2 modalities x 2 varieties x 3 repetitions x 12 informants). Data analysis covered six steps: coding of the repetitions; isolation; automatic segmentation in PRAAT; acoustic measurements; selection of the three best repetitions and application of the MatLab interface. The results from this relevant study revealed prosodic aspects in the speech from Belém and Macapá when indicated similar prosodic contours of F0 in the final stressed syllable in both modalities. However, prosodic features were more regular in yes-no question.

Keywords: Prosodic analysis. Sociophonetic. Brazilian portuguese. AMPER project.

1 Doutoranda do PPGL/UFPA. Email: rosinelelemos@gmail.com

2 Mestranda do PPGL/UFPA. Email: letcamila89@gmail.com

Introdução

Os estudos geoprosódicos nas perspectivas do projeto AMPER-POR na Amazônia iniciaram em 2007, na Universidade Federal do Pará (UFPA), coordenado pela professora Dra. Regina Cruz. O projeto em questão integra as pesquisas do AMPER que objetiva formar um grande Atlas Prosódico das Línguas Românicas.

Nesse contexto, a equipe do AMPER na Amazônia tem investigado o português falado na região e já conta com *corpora* formados em várias localidades paraenses. Atualmente, o projeto expandiu a sua investigação prosódica a outros Estados da Amazônia Legal, como Amapá³, Amazonas⁴ e Maranhão⁵.

Dentro desse viés de pesquisa, o objeto do presente estudo investiga a variação geoprosódica do português falado em duas capitais da Amazônia Oriental: Belém (PA) e Macapá (AP). Justifica-se a escolha dessas localidades, principalmente, por duas razões: a) as cidades selecionadas têm origens históricas comuns provenientes da formação do Estado do Grão Pará (1751); b) as localidades mapeadas compõem o banco de dados do projeto AMPER – POR: Brito (2014) e Lemos (em andamento).

Não obstante, este artigo tem como objetivo proceder a uma análise sociofonética interdialetoal do português falado em Belém e Macapá, para averiguar traços prosódicos semelhantes na sílaba tônica do sintagma nominal final em enunciados declarativos neutros e interrogativos totais, dos acentos oxítonos, como por exemplo, na frase “O Renato gosta do bisavô”. Em relação à comparação interdialetoal, foi utilizada a entoação modal como parâmetro de análise. Diante do propósito de investigar o padrão entoacional das duas variedades, as motivações para as hipóteses partiram do objeto deste estudo, que consiste em averiguar se as variedades alvo selecionadas possuem uma prosódia que caracterize o falar amazônico, considerando a formação sócio-histórica do Brasil Colonial. Neste sentido, a colonização portuguesa foi um fator que contribuiu fortemente na formação do português brasileiro em geral. Especificamente na Amazônia colonial, segundo Rodrigues (1996) a Língua Geral Amazônica (LGA) favoreceu a comunicação dos índios Tupi-Guarani com os portugueses à época da colonização. Assim, é possível que o português falado na Amazônia guarde algum traço prosódico daquele contato linguístico. Desse modo, o presente estudo encontra-se norteado pelas seguintes hipóteses:

3 Macapá - Lemos (em andamento), Oiapoque e Mazagão – Barros (em andamento).

4 Borba – Lemos (em andamento).

5 São Luís – Cardoso (em andamento).

a) Belém e Macapá têm semelhanças prosódicas por suas origens sócio-históricas;

b) o padrão entoacional do português falado em Belém e Macapá tem na região nuclear, isto é, no sintagma verbal (posição do objeto) da sentença declarativa neutra, a pretônica ascendente e a tônica descendente; a interrogativa total tem curvas descendentes na pretônica e ascendente na tônica, com contorno circunflexo, ou seja, esse padrão é igual ao descrito por Moraes (1993) para o português do Rio de Janeiro;

c) há traços prosódicos semelhantes entre as variedades de Belém e Macapá na última sílaba tônica do sintagma nominal final;

d) dentre os parâmetros físicos como a frequência fundamental (F0), a duração (medida em milissegundos) e a intensidade (medida em decibéis), a F0 é parâmetro mais relevante para atestar características prosódicas similares entre as variedades de Belém e Macapá, na sílaba tônica final dos enunciados declarativos neutros e interrogativos totais.

A fim de aceitar ou rejeitar as hipóteses elencadas, tomou-se como base os resultados de três trabalhos do AMPER-POR nesta comparação interdialetoal entre Belém e Macapá. Tais estudos serão descritos, a seguir:

No estudo de Lemos e Brito (2018) observaram-se traços prosódicos com contornos mais regulares nas sílabas tônicas das pautas acentuais oxítonas em relação às paroxítonas e proparoxítonas. Ainda no âmbito do AMPER-POR, Cruz, Seara e Moutinho (2015) constataram que existe similaridade prosódica possivelmente de origem açoriana entre duas variedades do português do Brasil: Belém e Florianópolis, em comparação com uma variedade do português europeu, a de São Miguel (Açores). A informação sobre a imigração açoriana será explicada na seção 3.2 deste artigo. Os resultados desse estudo foram relevantes neste artigo por dois motivos: a) Belém e Florianópolis têm origens históricas comuns, ligadas ao processo imigratório ao Brasil colonial. A esse respeito, de acordo com Chambouleyron (2010), os açorianos vieram ao Norte para impulsionar o povoamento na região; Furlan (1989), afirma que eles chegaram ao Sul; em Santa Catarina, com o mesmo objetivo, no século XVIII; b) há padrão prosódico de F0 semelhante na sílaba tônica final dos enunciados interrogativos totais.

A pesquisa de Nunes (2011) sob a metodologia AMPER revelou que a região nuclear da sentença é a mais importante para a caracterização dos dialetos catarinenses confrontados (Florianópolis e Lages – no Estado de Santa Catarina). Os resultados apontaram que é no núcleo da sentença que

se distinguem as declarativas neutras das interrogativas totais nos dois dialetos. Tais resultados são pertinentes com o estudo em questão, pois além de ter o mesmo foco de análise (núcleo da sentença) busca-se observar o comportamento prosódico das frases oxítonas, uma vez que no estudo da autora, as declarativas neutras tiveram picos de F0 mais proeminentes do que das interrogativas totais.

Neste estudo sociofonético, investiga-se a variação prosódica da fala por meio de uma análise acústica da F0. Todavia, os fatores sociais (sexo e idade) dos informantes podem ser considerados em uma análise, mas nesta comparação interdialetoal não houve diferenças relevantes entre essas variáveis. Como suporte teórico, adotou-se a Sociofonética (THOMAS, 2011) que estuda as variações da fala em diferentes contextos sociais.

Por fim, a metodologia AMPER-POR para a análise das sentenças declarativas neutras e interrogativas totais possibilitou chegar aos resultados aqui empreendidos que serão discutidos com outros estudos prosódicos do PB que investigaram a F0 em sentenças declarativas neutras e interrogativas totais. A seguir, serão apresentadas as seções que nortearam a organização deste estudo.

A seção 1 é dedicada à introdução, na seção 2, apresenta-se a Sociofonética como embasamento teórico adotado, na seção 3 abordam-se os estudos prosódicos que favoreceram essa comparação interdialetoal, assim como a descrição de Moraes (1993) para o Rio de Janeiro; a descrição de Nunes (2011); as descrições de Silva (2011) e Silvestre (2012) com foco em Belém e Macapá. A seção 4 é dedicada à metodologia AMPER; na seção 5, mostram-se os resultados da análise acústica empreendida sobre a comparação geoprosódica entre Belém (PA) e Macapá (AP) e a discussão dos seus resultados, na 6, as conclusões e, finalmente, as referências.

2 Arcabouço teórico: a Sociofonética

A Sociofonética estuda a fala no contexto social e as variações linguísticas provenientes da produção e percepção dos falantes. Seu vasto campo de investigação contempla tanto as variáveis qualitativas quanto as quantitativas. No estudo em questão, adotou-se essa teoria para analisar a variação fonética (acústica) e as variáveis sociais como sexo e escolaridade dos informantes.

Thomas (2011) assegura que a Sociofonética mantém uma interface entre a Fonética e a Sociolinguística, áreas da Linguística que “(...) convergem para a meta de compreender os aspectos cognitivo e diacrônico da linguagem⁶”. (THOMAS, 2011, p. 2). De acordo com o autor, historicamente, as línguas variam e o fenômeno da variação é uma propriedade fundamental da linguagem. Assim,

6 “(...) they actually converge on the goal of understanding the cognitive and diachronic aspects of language”. (Tradução nossa).

devido ao aspecto diacrônico da variação e mudança linguística no tempo, no que concerne às variações fonéticas e sociais, a Sociolinguística compartilha com a Fonética a descrição e o entendimento da mudança sonora no tempo.

Entretanto, os fundamentos do que hoje é chamado de Sociofonética foram estabelecidos por Labov (1966) em seus estudos sobre a estratificação social do inglês na cidade de Nova York.

Os primeiros estudos sociofonéticos focaram suas análises nas vogais. Atualmente, o seu campo de atuação compreende a investigação e a análise instrumental de outros tipos de sons da fala. Todavia, neste artigo, a Sociofonética é essencial à compreensão da variação social e prosódica do português amazônico. A análise do parâmetro acústico da (F0) tem como foco as vogais das sentenças selecionadas. Baranowski (2013) afirma que as vogais são o centro de interesse das análises acústicas, uma vez que nelas concentram-se mais energia acústica produzida por falantes numa amostra de fala.

A primeira pesquisa intitulada como Sociofonética foi realizada pelo pesquisador Deshaies-Lafontaine (1974), sobre a variação do francês canadense, no estudo de sete variáveis fonéticas /a/, /e/, /eu/, /o/, /g/, /ch/ e /r/. O termo, usado como paralelo a “sociolinguístico”, centraliza sua investigação nas variáveis fonéticas, em detrimento às sintáticas ou lexicais. Embora os estudos em Sociofonética ainda sejam incipientes, muitos linguistas têm voltado suas investigações não somente à produção da fala, mas também buscam entender como a variação dos sons é percebida pelo ouvido humano durante as comunicações. De fato, a Sociofonética é um campo vasto de investigação científica que engloba o domínio fonético, fonológico, lexical, sintático-semântico, e mais recentemente, da Fonética Forense.

Por tratar-se de um estudo fonético e social, justifica-se aqui essa teoria, uma vez que os locutores de Belém e Macapá produziram sentenças declarativas neutras e interrogativas totais, estudadas do ponto de vista acústico (F0) e pelas variações sociais (sexo e escolaridade). A próxima seção contempla algumas descrições prosódicas da entoação no PB e os estudos que motivaram esta pesquisa.

3 Estudos sobre a entoação modal no português do Brasil e no AMPER-POR

Os estudos sobre a entoação modal no português no PB têm sido objeto de investigação suprasegmental com a finalidade de descrever as variações geoprosódicas em contextos de produção da fala. A entoação, além de abarcar as análises dos parâmetros de frequência fundamental (F0), da duração (medida em ms) e da intensidade (medida em dB), discrimina os tons de descida e de subida

das curvas de F0, assim como determina os padrões acentuais de determinada língua. Desse modo, a entoação indica o tipo de acento e modalidade frasal. Ela sinaliza os aspectos linguísticos, como por exemplo, se um enunciado é declarativo ou interrogativo e, ainda, expressa atitude e emoção do falante. Nas subseções, a seguir, serão apresentados alguns estudos sobre a entoação modal no PB, assim como as descrições prosódicas no âmbito do AMPER-POR que motivaram este estudo.

3.1 A entoação modal no português do Brasil (PB)

A investigação sobre a variação prosódica no PB tem sido objeto de estudo de muitos pesquisadores empenhados em descrever os aspectos suprasegmentais da fala como: o acento, Cruz et al. (2017), o ritmo (BARBOSA, 2012) e a entoação (CUNHA, 2000). Tais trabalhos foram bastante relevantes para os estudos em prosódia no PB. Outras pesquisas nessa área, já estabeleceram padrões entoacionais de diversas variedades linguísticas do PB. Desse modo, com o propósito de corroborar com este estudo, optou-se por referenciar três descrições do PB, que analisaram a F0 e a região nuclear das sentenças declarativas neutras e interrogativas totais. Assim, os trabalhos revisados foram os de Moraes (1993) que estabeleceu um padrão entoacional para o português carioca e os de Silva (2011) e Silvestre (2012) que analisaram entre o de outras capitais, o padrão entoacional de Belém e Macapá.

Moraes (1993) descreveu diferentes modalidades de entoação modal na variedade do português falado no Rio de Janeiro, com o objetivo de investigar a configuração geral da curva de F0, que segundo o autor, é o parâmetro mais importante para a entoação modal. Ele analisou diversos tipos de enunciados, porém, neste estudo, observaram-se apenas os assertivos e interrogativos em posição final do enunciado produzidos por falantes da variedade culta carioca. Os resultados apontaram os seguintes padrões melódicos: o declarativo apresentou ataque em um nível médio baixo, com uma ligeira ascensão, seguido de uma pretônica baixa, que continua baixa até a tônica final de todo enunciado. O interrogativo total exibiu uma pretônica baixa, seguido de uma tônica final bastante elevada, com queda na postônica, configurando um contorno circunflexo, com alinhamento de pico à direita da tônica e os níveis baixos agregam-se às átonas finais.

Silva (2011) descreveu a prosódia de enunciados interrogativos totais nos falares de 25 capitais brasileiras⁷. Cabe aqui observar o comportamento prosódico de Belém e Macapá. Nesse estudo, o

7 Macapá, Boa Vista, Manaus, Belém, Rio Branco, Porto Velho, São Luís, Teresina, Fortaleza, Natal, João Pessoa, Recife, Maceió, Aracaju, Salvador, Cuiabá, Campo Grande, Goiânia, Belo Horizonte, São Paulo, Vitória, Rio de Janeiro, Curitiba, Florianópolis e Porto Alegre.

parâmetro acústico controlado foi a frequência fundamental, com o objetivo de observar as marcas de variação regional nos acentos pré-nucleares e nucleares dessas sentenças do PB. O *corpus* da pesquisa foi composto por 200 dados do projeto *Atlas Linguístico do Brasil* (ALiB), sendo que, para cada capital foram selecionados quatro informantes de duas faixas etárias: a primeira de 18 a 30 anos e a segunda de 50 a 65 anos. Os dados foram segmentados no programa PRAAT com a finalidade de medir os valores de F0 nas sílabas.

O estudo da autora revelou dois padrões melódicos para a questão interrogativa total na variedade de Belém na região nuclear das sentenças. No primeiro, a frequência aumenta na tônica final atingindo seu ápice na segunda metade dessa sílaba, as postônicas apresentaram movimento descendente. No segundo, há um pico de F0 mais elevado na sílaba tônica final e apagamento das postônicas finais.

No que se refere à variedade de Macapá, os resultados do estudo mostraram um único padrão melódico para as interrogativas totais: proeminência de F0 na primeira sílaba tônica, acompanhada de um movimento descendente até a última pretônica. A partir dessa sílaba, verifica-se outra proeminência melódica, caracterizada por um movimento ascendente com pico à direita da sílaba tônica, porém com altura inferior ao primeiro pico. O comportamento da F0 não pôde ser analisado na postônica final por haver desvozeamento em todos os dados de Macapá.

Em síntese, dentre os dois padrões prosódicos das interrogativas totais descritos por Silva (2011) para a variedade de Belém, apenas o primeiro (ascendente/descendente) é semelhante ao circunflexo, descrito por Moraes (1993) para o Rio de Janeiro; o segundo, é ascendente, com pico de F0 na tônica, porém apresenta apagamento nas postônicas, o que diverge do padrão carioca. Por outro lado, Macapá registrou apenas um padrão prosódico ascendente na tônica, esse comportamento também diverge daquele encontrado para o Rio de Janeiro. Comparando Belém e Macapá, pode-se dizer que ambas têm maior pico de F0 nas tônicas nos enunciados interrogativos totais.

Seguindo a investigação do português brasileiro, Silvestre (2012) descreveu a entoação regional em sentenças declarativas neutras nos dialetos de 25 (vinte e cinco) capitais brasileiras com o objetivo de estabelecer um padrão prosódico para cada capital selecionada, com dados e metodologia do projeto ALiB, descrita anteriormente. O parâmetro acústico controlado foi a frequência fundamental e o foco da análise incidiu especialmente nos acentos pré-nuclear e nuclear de enunciados declarativos. Os dados foram tratados no programa PRAAT com a finalidade de segmentar e medir os valores de F0 nas sentenças.

Na análise em questão, a pesquisadora encontrou dois padrões para variedade de Belém e dois padrões para Macapá. Nos dados de Belém, foram observados dois padrões melódicos diferentes, com relação ao comportamento da F0, no contorno nuclear para a declarativa neutra. O primeiro, apresenta a seguinte configuração: tom alto nas sílabas pretônicas e movimento decrescente, que se inicia na última sílaba tônica final do enunciado e continua na postônica. O segundo padrão verificado na capital paraense exhibe tons altos nas pretônicas iniciais, com movimento descendente nas tônicas que não chega até a sílaba postônica final, o que configura uma ascensão na postônica recorrente na fala feminina.

Com relação a Macapá, os resultados de Silvestre (2012) apontaram dois padrões melódicos para as declarativas neutras: proeminência inicial da F0 na primeira sílaba tônica nos dois padrões melódicos. Já as sílabas postônicas finais apresentaram em dois dados femininos da segunda faixa etária uma alternância em relação à sílaba tônica final do enunciado: ora apresenta uma configuração ascendente, ora descendente.

Por conseguinte, os dois diferentes padrões entoacionais das declarativas neutras para Belém têm as tônicas descendentes, o que se assemelha com o padrão observado por Moraes (1993). Macapá também apresentou nos dois padrões, descendência de F0 na sílaba tônica, portanto, está em consonância com o descrito para o Rio. O comportamento das postônicas apresentou variação de subida e descida, ou seja, divergem do carioca. Entretanto, a proeminência de F0 acontece na tônica nas duas variedades.

Esses estudos prosódicos acima descritos apresentaram os padrões encontrados em pesquisas relevantes sobre a prosódia de sentenças declarativas neutras e interrogativas totais no PB.

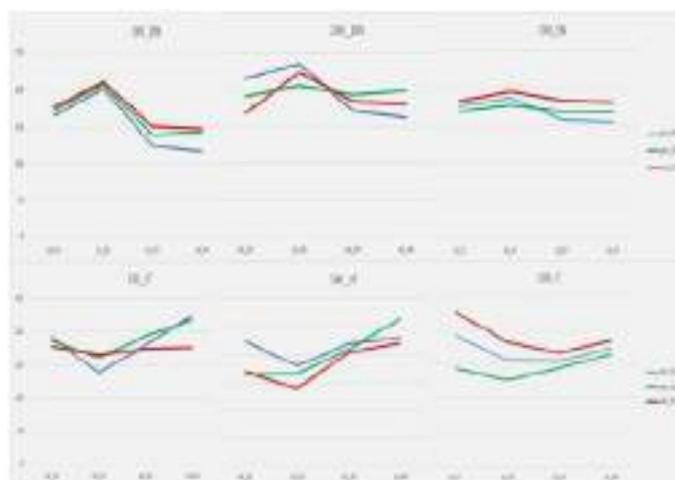
3.2 A entoação modal no âmbito do projeto AMPER-POR

As pesquisas realizadas pelo projeto AMPER-POR no Brasil têm contribuído com a investigação prosódica de diversas variedades do PB. Dentre essas, destaca-se o estudo de Cruz, Seara e Moutinho (2015) e Nunes (2011), cujos resultados incentivaram a investigação aqui empreendida. Com relação à equipe AMPER-Amazônia, uma descrição prosódica foi bastante motivadora para confrontar as variedades de Belém e Macapá.

Lemos e Brito (2018) ao realizarem uma análise comparativa entre as variedades de Belém, Macapá e Borba (AM), especificamente no Sintagma Nominal Final (doravante SNF) das sentenças declarativas neutras e interrogativas totais na pauta acentual oxítônica, observaram que o padrão

entoacional das sentenças declarativas neutras apresentaram as curvas de F0 altas desde a preposição (de), com maior elevação nas pretônicas e queda de frequência na última sílaba tônica do enunciado. Nesse ponto, há uma semelhança prosódica entre os dialetos, o que demonstra maior regularidade no comportamento entoacional das curvas de F0. No que tange às sentenças interrogativas totais a F0 caracterizou-se com contorno entoacional mais baixo de frequência na preposição (de) que se estendeu às pretônicas e, por fim, com pico entoacional elevado na última sílaba tônica, onde as curvas se aproximaram, configurando traços melódicos semelhantes. Por outro lado, as pautas paroxítonas e proparoxítonas, apesar de realizarem pico entoacional na última sílaba tônica dos enunciados, não apresentaram contornos melódicos bem definidos. O gráfico abaixo apresenta a configuração prosódica observada na pauta acentual oxítona.

Gráfico 1- Contorno melódico das sentenças declarativas neutras (gráficos acima) e interrogativas totais (gráficos abaixo).



Legenda: Gráficos de F0 comparativos entre as falas femininas de Belém (linha azul), Macapá (linha vermelha) e Borba (linha verde) das frases oxítonas bisavô, em posição nuclear.

Fonte: Lemos e Brito (2018).

Ainda no âmbito do projeto AMPER-POR, Cruz, Seara e Moutinho (2015) confrontaram três variedades do português com objetivo muito similar ao do presente estudo, ao realizarem uma análise interdialetoal entre Belém (PA), Florianópolis (SC) e São Miguel (Açores). A escolha dessas duas localidades do Norte e Sul do Brasil está relacionada à imigração açoriana à época da colonização portuguesa no país. Segundo registros históricos de Furlan (1989), o Norte e o Sul do Brasil faziam parte da estratégia política do governo português sob o controle da colônia, por isso a Coroa portuguesa incentivava a vinda de casais açorianos com o seguinte propósito:

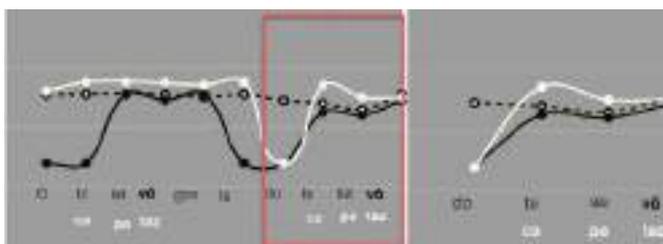
Diminuir a sobrecarga demográfica dos Açores e para tomar posse dos extremos Norte e Sul do Brasil, expostos à invasão espanhola, que a Coroa portuguesa estimulou e dirigiu de 1748/1756 a vinda de açorianos e madeirenses ao Norte (Maranhão e Grão Pará) e ao Sul - hoje Rio Grande do Sul e Santa Catarina. (FURLAN, 1989, p. 83).

Os resultados desse estudo preliminar mostraram similaridade prosódica entre os três dialetos e, a partir desse confronto, as autoras chegaram às seguintes conclusões:

i. Há traços prosódicos semelhantes possivelmente de origem açoriana entre as variedades de Belém e Florianópolis em comparação com o Açores, no contexto da sílaba tônica final;

ii. As curvas de F0 assumem características prosódicas similares na posição nuclear das sentenças, conforme apresenta a figura 1, a seguir:

Figura 1 – Contorno melódico das curvas de f0 de Belém (PA), Florianópolis (SC) e Açores (PT) da modalidade interrogativa total, dos acentos oxítonos: capataz (PE) e bisavô (PB).



Legenda: F0 comparativa entre as falas de Belém (linha preta contínua), Florianópolis (linha preta pontilhada) e Açores (linha branca) em frases com vocábulos oxítonos em posição nuclear.

Fonte: Cruz, Seara e Moutinho (2015, p.12).

Observando os resultados desse estudo conclui-se que as curvas entoacionais oxítonas **bisavô** (PB) e **capataz** (PE) estão na mesma posição na última sílaba tônica, o que evidencia proximidade prosódica entre Belém, Florianópolis e São Miguel (Açores).

Consoante também à metodologia AMPER, Nunes (2011) descreveu o padrão entoacional das sentenças declarativas neutras e interrogativas totais nos falares de duas cidades catarinenses: Florianópolis e Lages. A autora analisou as sentenças tanto na região pré – nuclear (posição sintática do sujeito) quanto na região nuclear (posição sintática do objeto) com sintagmas nominais simples e estendidos, encontrando, assim, padrões de curvas entoacionais distintos, dependendo do tipo de sintagma e da sua posição sintática para cada dialeto catarinense analisado.

Com relação ao sintagma nominal na posição de objeto, Nunes (2011) analisou a sílaba tônica nuclear. No sintagma nominal final, as curvas melódicas das declarativas, de uma maneira geral, exibiram picos mais proeminentes do que as das interrogativas, com elevação da F0 alinhada à sílaba pretônica e sua queda ocorrendo na sílaba tônica, padrão este encontrado para o português falado no Sul do Brasil, conforme observado na figura 2.

Figura 2 – Padrões declarativos e interrogativos de Florianópolis e Lages.

	oxítonas	paroxítonas	proparoxítonas
Masc Fpolis sem extensão	↑ ↑ ↑	× × ×	× × ×
Masc Lages sem extensão	↘ ↘ ↘	↘ ↘ ↘	↘ ↘ ↘
Masc Fpolis c/ extensão Adj	↑ ↘ ↘	↑ ↘ ↘	↘ ↘ ↘
Masc lages c/ extensão Adj	↘ ↘ ↘	↘ ↘ ↘	↘ ↘ ↘
Masc Fpolis c/ extensão S.Prej	↑	↘	↘
Masc Lages c/ extensão S.Prej	↘	↘	↘

Legenda: A cor vermelha refere-se às sentenças declarativas e a azul às interrogativas.

Fonte: Nunes (2011, p.112).

Os resultados de Nunes (2011) para as três pautas acentuais mostraram um padrão entoacional distintivo nas duas modalidades frasais analisadas. Porém, essa diferença no movimento intrassilábico é mais evidente na pauta acentual paroxítona em relação às oxítonas e proparoxítonas.

4 Metodologia AMPER

Nesta seção, apresenta-se a metodologia do projeto AMPER, enfatizando o AMPER-POR, o AMPER-Amazônia e a formação do *corpus* de Belém e Macapá.

O projeto AMPER, criado em 1999 por pesquisadores do Centro de Dialectologia da Universidade de Grenoble (França), impulsionou os estudos geoprosódicos ao estabelecer uma metodologia única, com *corpora* comparáveis no espaço dialetal românico. Isto possibilita, portanto, uma comparação interdialetoal e intradialetoal entre as variedades do português europeu e do brasileiro.

No sentido de viabilizar pesquisas dialetais comparáveis para todas as línguas românicas, o AMPER desenvolveu uma metodologia padronizada, com mesma estratégia de inquérito, mesmo procedimento de análise instrumental multiparamétrico - um Questionário Comum de Base (QCB) composto de sentenças produzidas em duas modalidades entoacionais (declarativa neutra e interrogativa total). Quanto ao tipo de acento, o *corpus* AMPER contempla os vocábulos-alvo nas três pautas acentuais do português: proparoxítona, paroxítona e oxítona, assim como, na seleção dos sujeitos da pesquisa, controla as variáveis sociais como escolaridade (fundamental, médio e superior), sexo (feminino e masculino), faixa etária dos informantes (acima dos 30 anos), devendo estes serem, obrigatoriamente, nativos de cada localidade pesquisada.

O projeto possui ainda mecanismos de análise acústica e gráficos representativos comuns a todas as línguas alvo e suas variedades. Para a análise acústica, o AMPER utiliza como recurso dois *softwares*: a) o *MatLab*, versão R2016b, com scripts criado por Antonio Romano (1999), no qual utilizam-se a *Interface AMPER_beta11* criada por Albert Rilliard e b) o programa *PRAAT*, versão 5.1.10, criado por Boersma e Wennink (2011).

Em consonância com o objetivo do AMPER de formar um grande atlas prosódico das línguas Românicas, e de acordo com o propósito metodológico do projeto descrito anteriormente, é possível comparar as diferentes variedades linguísticas que já dispõem de descrições prévias no banco de dados AMPER, utilizadas em todos os estudos comparativos no âmbito do projeto.

Com o intuito de mapear a variação prosódica do português (PE e PB) foi criado o subprojeto AMPER-POR, coordenado por Lurdes de Castro Moutinho, na Universidade de Aveiro, em Portugal. O projeto disponibiliza sua base de dados online, no site <<http://www.varialing.eu/>>. Nesse endereço eletrônico, estão todas as informações sobre o projeto AMPER-POR e os links de cada um dos outros domínios linguísticos inerentes ao AMPER.

4.1 AMPER equipe Amazônia

O AMPER-Amazônia é coordenado pela prof^a. Dr^a. Regina Cruz na UFPA, com o objetivo de mapear o nível prosódico do português brasileiro falado na Amazônia. O projeto possui *corpora* formados nos estados do Pará, Amapá, Amazonas e Maranhão, conforme pode ser observado no mapa 1, a seguir:

Mapa 1 – Localidades com *corpora* formados ou em formação pelo Projeto AMPER na Amazônia



Fonte: Fernandes (em andamento).

O estudo em questão está diretamente vinculado ao AMPER-Amazônia, cujos dados compõem a investigação prosódica de Macapá Lemos (em andamento) e de Belém (BRITO, 2014). Na subseção a seguir, apresenta-se a formação do *corpus* de Belém e Macapá.

4.1.1 Sujeitos da pesquisa (Belém e Macapá)

Participaram da pesquisa 12 informantes nativos, sendo seis de Belém e seis de Macapá. Para cada variedade alvo, selecionaram-se três locutores femininos e três masculinos, que pertencem aos níveis de escolaridade fundamental, médio e superior. A estratificação social segue a metodologia empreendida no projeto AMPER, codificando os participantes da pesquisa conforme é observado no quadro 1.

Quadro 1 – Codificação e estratificação social dos locutores de Belém (BE0) e Macapá (BG1).

Variedade	Locutor	Sexo	Escolaridade	Duração do sinal gravado
Macapá	BG11	Feminino	Ensino Fundamental	1h:30 min
	BG12	Masculino	Ensino Fundamental	1h:41 min
	BG13	Feminino	Ensino Médio	54 min
	BG14	Masculino	Ensino Médio	1h:30 min
	BG15	Feminino	Ensino Superior	1h: 06 min
	BG16	Masculino	Ensino Superior	1h:10 min
Belém	BE01	Feminino	Ensino Fundamental	54 min 24s
	BE02	Masculino	Ensino Fundamental	38 min 31s
	BE03	Feminino	Ensino Médio	40 min 19s
	BE04	Masculino	Ensino Médio	39 min 47s
	BE05	Feminino	Ensino Superior	31 min 46s
	BE06	Masculino	Ensino Superior	30 min 13s

Fonte: elaboração própria.

4.1.2 *Corpus*

O *corpus* do AMPER-POR que investiga o PB é formado por 66⁸ sentenças, assim como o estendido, com 102⁹ frases, que obedecem as mesmas fonéticas e sintáticas. Dessa forma, as frases são do tipo SVC (sujeito + verbo + complemento) com a inclusão de sintagmas adjetivais e preposicionados. Todas as frases possuem: 1) três personagens: Renato, pássaro e bisavô; 2) três sintagmas adjetivais:

8 O *corpus* de Belém foi o simples, contendo 66 frases.

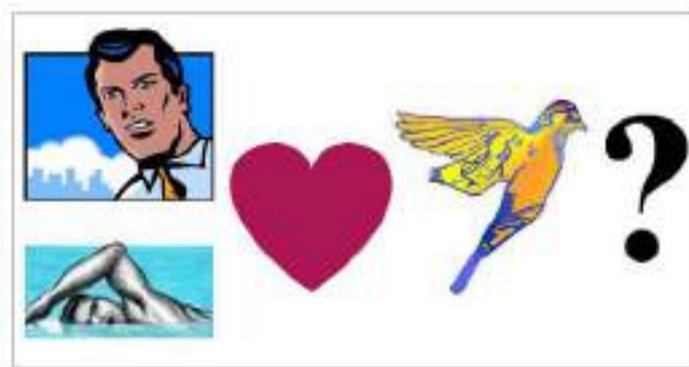
9 O *corpus* de Macapá foi o expandido, contendo 102 frases. No *corpus* expandido foram introduzidas sentenças com o vocábulo *capataz*.

nadador, bêbado e pateta; 3) três sintagmas preposicionados indicadores de lugar: de Mônaco, de Veneza e de Salvador; 4) um único verbo: gostar.

Os *corpora* de Belém e Macapá foram coletados com um gravador digital profissional *Marantz* modelo PMD660 e um microfone *Shure* de cabeça dinâmico para a melhor captura do áudio. As configurações do gravador foram especificadas com a taxa de amostragem em 44.100 Hz, 16 bits, modo mono, formato *.wav*; para preservar a qualidade acústica do sinal sonoro.

A gravação dos sinais sonoros foi feita através de uma leitura semi-espontânea de imagens, combinadas para a formação das frases. Ressalta-se que os informantes não tiveram contato com a frase escrita. A figura 3, a seguir, apresenta um exemplo de uma sentença formada por estímulo visual.

Figura 3 – Exemplo de estímulo visual para produção de frases interrogativas totais: O Renato nadador gosta do pássaro?



Fonte: elaboração própria.

Concluída a coleta, o tratamento dos dados compreendeu sete fases, dos quais, os cinco primeiros foram realizados com o auxílio do software PRAAT versão 5.1.10. As etapas do trabalho foram: i) codificação das repetições; ii) isolamento dos sinais sonoros em arquivos individuais; iii) segmentação automática (realizada pelo *script lance_batch_easyalign_V3.praat* versão 5.1.10, criado por Albert Rilliard); iv) extração das medidas acústicas dos segmentos vocálicos e das médias dos parâmetros físicos controlados; v) seleção das 3 melhores repetições; vi) aplicação da Interface Matlab e vii) elaboração de gráficos no software R, versão 4.0.0 (2020-04-24).

Dessa forma, todos os *corpora* AMPER adotam esses procedimentos metodológicos em suas pesquisas. Por isso, além de Belém e Macapá, outro estudo a utilizar essa metodologia foi o trabalho de Lemos e Cruz (2016), que descreveram a variação prosódica do português falado em Baião (PA).

As autoras supracitadas investigaram a entoação modal de sentenças declarativas neutras e interrogativas totais da variedade baionense, com o objetivo de observar o papel da F0 na discriminação

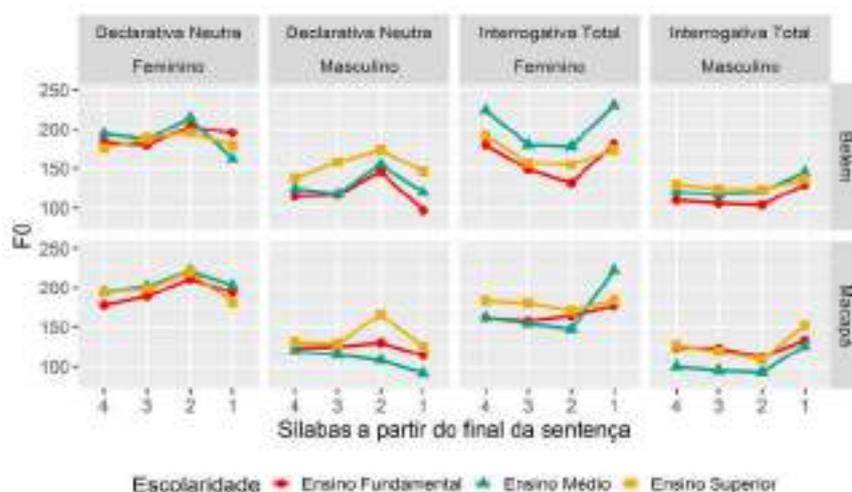
das duas modalidades frasais. Para a análise acústica, selecionaram-se amostras de fala de dois locutores femininos e dois masculinos, moradores nativos da localidade alvo, com faixa etária acima de trinta anos, dos níveis de escolaridade fundamental e médio. Analisaram-se vinte e uma sentenças do *corpus* AMPER contendo sintagmas nominais finais simples e estendidos nas três pautas acentuais do português. Os resultados apontaram que a F0 é determinante na distinção da entoação modal, com as principais variações de frequência ocorrendo na última sílaba tônica do sintagma nominal final dos enunciados. Portanto, nas descrições AMPER Amazônia, a F0 tem se revelado um parâmetro importante tanto na distinção das modalidades frasais quanto na comparação interdialetoal, que é o parâmetro investigado entre Belém e Macapá neste artigo.

5 Análise interdialetoal entre as variedades de Belém (PA) e Macapá (AP)

Nesta seção, apresentam-se os resultados da comparação interdialetoal entre as duas variedades amazônicas, por meio de análises acústicas da F0, em sentenças declarativas neutras e interrogativas totais dos sintagmas nominais finais da pauta acentual oxítone. Essas análises foram feitas com base nas medidas acústicas dos arquivos fonos gerados pela interface *Matlab*. Esses arquivos contêm os valores de F0 das três frases oxítonas selecionadas para a análise.

Para este estudo, as três sentenças declarativas neutras e as três interrogativas totais selecionadas foram : **pwk** (O pássaro gosta do bisavô), **twk** (O Renato gosta do bisavô) e **kwk** (O bisavô gosta do bisavô). Os gráficos a seguir mostram os resultados das análises referentes ao sintagma nominal final das sentenças supracitadas, representado pelo vocábulo **bisavô**. Observar o conjunto de gráficos 2.

Gráfico 2 – Contorno melódico da sentença **pwk** (O pássaro gosta do bisavô), nas modalidades declarativas neutras e interrogativas totais.



Legenda: Gráficos de F0 comparativos entre as falas de Belém e Macapá, dos informantes do sexo feminino e masculino em posição nuclear da sentença. As linhas vermelhas representam os informantes do ensino fundamental

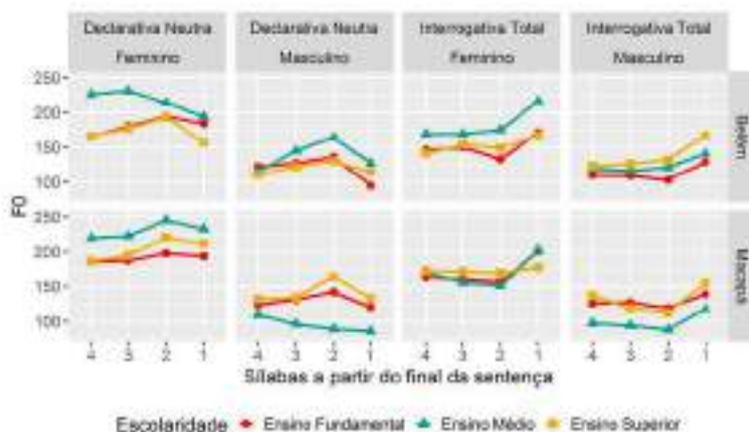
(EF), as azuis correspondem aos informantes do ensino médio (EM) e as laranjas, indicam o ensino superior (ES).

Fonte: elaboração própria.

Os gráficos acima apresentam a configuração melódica da sentença **pwk** (O pássaro gosta do bisavô) de sentenças declarativas neutras e interrogativas totais, dos informantes de Belém e Macapá. A contagem das sílabas, na ordem de 4 a 1, significa (s4=do; s3=bis; s2=sa; s1=vô). As declarativas neutras nos dados femininos e masculinos, nas duas variedades apresentaram o seguinte padrão entoacional: contornos iniciais elevados na sílaba (a partir de agora s) (s4) que continuam assim até as segundas pretônicas (s2), em que ocorrem os picos de F0; posteriormente, apresentam uma queda que vai até às tônicas (s1), formando assim o movimento circunflexo; nestes pontos (s1), nas últimas sílabas tônicas, pode-se observar que os contornos melódicos das duas variedades são semelhantes, no (SNF). As interrogativas totais para os informantes femininos e masculinos das duas variedades, somente apresentaram contornos melódicos descendentes nas (s4) até às pretônicas (s3 e s2), com ascendência nas últimas (s1), em que ocorrem os picos de F0. Embora os informantes femininos do ensino médio das duas variedades apresentarem contornos melódicos mais altos, eles acontecem na última sílaba tônica do enunciado, o que evidencia similaridade prosódica para o padrão investigado: descendente nas declarativas neutras e ascendentes nas interrogativas totais.

Com relação à sentença **twk** (O Renato gosta do bisavô) os falantes femininos e masculinos de Belém e Macapá mostraram que as declarativas neutras têm pico de F0 nas pretônicas e queda de frequência nas últimas sílabas tônicas das (s1), com exceção para o informante do sexo masculino do ensino médio de Macapá, que realizou queda desde as (s4, s3, s2) atingindo maior declínio nas (s1). Esse desvio talvez seja aleatório, já que ocorreu apenas nessa frase. Os informantes das interrogativas totais, nas duas variedades e modalidades frasais produziram contornos melódicos baixos nas (s4, s3 e s2) com pico de F0 nas (s1), o que configura similaridade prosódica, conforme os gráficos 3.

Gráfico 3 – Contorno melódico da sentença **twk** (O Renato gosta do bisavô) nas modalidades declarativas neutras e interrogativas totais.



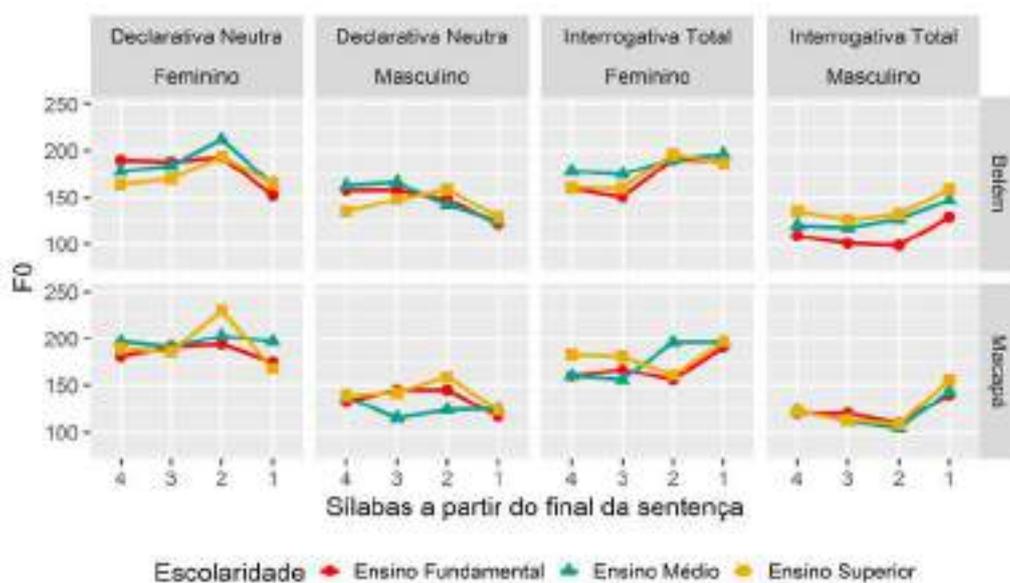
Legenda: Gráficos de F0 comparativos entre as falas de Belém e Macapá, dos informantes do sexo feminino e

masculinos em posição nuclear da sentença. As linhas vermelhas representam os informantes do ensino fundamental (EF), as azuis correspondem aos informantes do ensino médio (EM) e as laranjas, indicam o ensino superior (ES).

Fonte: elaboração própria.

Quanto à sentença **kwk** (O bisavô gosta do bisavô) para as duas variedades e sexo (feminino) as declarativas apresentaram contorno circunflexo, que se inicia baixo nas (s4, s3) com elevação do pico nas últimas pretônicas (s2) e queda nas tônicas (s1). As interrogativas totais dos informantes masculinos de ambas as variedades apresentaram contorno ascendente com maior similaridade em relação às declarativas. Entretanto, os informantes femininos do ensino médio de Macapá e do fundamental e superior de Belém, não realizaram o comportamento ascendente nas tônicas. Verificar no conjunto dos gráficos 4.

Gráfico 4 – Contorno melódico da sentença **kwk**, (O bisavô gosta do bisavô) nas modalidades declarativas neutras e interrogativas totais.



Legenda: Gráficos de F0 comparativos entre as falas de Belém e Macapá, dos informantes do sexo feminino e masculinos em posição nuclear da sentença. As linhas vermelhas representam os informantes do ensino fundamental (EF), as azuis correspondem aos informantes do ensino médio (EM) e as laranjas, indicam o ensino superior (ES).

Fonte: elaboração própria.

5.1 Discussão dos resultados

Os resultados da comparação interdialetoal entre Belém e Macapá, em linhas gerais, mostraram que há traços prosódicos semelhantes na última sílaba tônica dos vocábulos oxítonos, tal similaridade acontece no contorno melódico das variedades confrontadas. Observou-se que esse contorno de F0 similar é mais regular principalmente na modalidade interrogativa total.

Os resultados do estudo de Cruz, Seara e Moutinho (2015), que analisaram sentenças interrogativas totais, observou-se que na sílaba tônica houve contorno de F0 semelhante nas duas modalidades, na pauta acentual oxítona, na sílaba tônica final do enunciado, o que evidenciou similaridades prosódicas possivelmente de origem açoriana entre Belém, Florianópolis e São Miguel. Neste estudo comparativo entre Belém e Macapá, o padrão entoacional observado nas duas modalidades apontaram que na região nuclear das sentenças, o contorno de F0 é bastante aproximado, principalmente nas interrogativas totais, as quais demonstraram curvas ascendentes regulares no SNF, que é o foco da análise. Neste aspecto, os resultados deste estudo estão em consonância com os das autoras. No que se refere à influência da colonização portuguesa nos falares amazônicos é pertinente uma investigação mais aprofundada para comprovar se a semelhança prosódica se justifica pelo contexto da formação sócio-histórica de Belém e Macapá.

No que tange ao padrão entoacional da modalidade interrogativa total do português falado em Belém, Silva (2011) encontrou um padrão para as interrogativas totais no sintagma nominal final dos enunciados: pico de F0 alinhando-se à direita da sílaba tônica e os níveis mais baixos associando-se às sílabas átonas adjacentes, tais caracterizações de F0 estão em consonância com a variedade de Belém, aqui analisada.

Quanto a Macapá, a autora observou um único padrão para as interrogativas totais: proeminência de F0 na sílaba tônica inicial, acompanhada de um movimento descendente até a última pretônica. A partir dessa sílaba, há um movimento melódico ascendente com pico à direita da sílaba tônica. Os resultados da autora acompanharam o contorno melódico ascendente na última sílaba tônica das interrogativas totais, observado neste estudo na variedade de Macapá.

Ao investigar o padrão entoacional da modalidade declarativa neutra, Silvestre (2012) encontrou dois padrões para Belém. O primeiro apresenta F0 alto no acento pré-nuclear, com configuração descendente, que se inicia na última sílaba tônica final do enunciado e continua na postônica. O segundo mostra uma proeminência da F0 na tônica inicial e tons altos nas sílabas adjacentes seguido de uma queda da F0 na última sílaba tônica. Quanto a Macapá, a autora encontrou dois padrões entoacionais, ambos com elevação da F0 na primeira sílaba tônica. As tônicas finais do enunciado se alternaram, ora com contorno descendente da tônica para a postônica, ora com maior elevação na postônica em relação à tônica. Esses resultados encontrados pela autora não estão em consonância com os achados neste estudo, pois na posição final do enunciado, a sílaba tônica apresenta sempre um contorno descendente.

O estudo de Nunes (2011) na investigação do português catarinense revelou que no sintagma nominal na posição de objeto, as declarativas mostraram pico entoacional mais proeminentes nas declarativas oxítonas do que nas interrogativas totais. Nesse aspecto, os resultados das análises de Belém e Macapá também apresentaram picos de F0 mais altos nas declarativas oxítonas em relação às interrogativas totais. Para o Rio de Janeiro, descrito por Moraes (1993) as interrogativas, no SNF têm proeminência acentual nas pretônicas e queda nas tônicas finais, o que configura o movimento circunflexo. Nos resultados de Belém e Macapá não foram observados movimento circunflexo nas interrogativas totais. Nas declarativas houve movimento circunflexo, com contornos ascendentes nas pretônicas e descendentes nas tônicas oxítonas.

Conclusões

O presente estudo compreendeu uma comparação geoprosódica interdialetoal entre as variedades de Belém e Macapá, com o objetivo de observar semelhança prosódica no contorno das curvas de F0, em sílabas tônicas finais de vocábulos declarativos neutros e interrogativos totais, da pauta acentual oxítona. As análises das variáveis fonéticas possibilitaram caracterizar a prosódia das duas cidades e, a partir da observação visual dos resultados, conclui-se que há semelhança na produção dos falantes belenenses e macapaenses.

Com relação às variáveis sociais, os informantes foram estratificados em sexo, escolaridade e faixa etária. No entanto, possivelmente por questões metodológicas como: amostra pequena, análise de apenas uma pauta acentual e um parâmetro físico, as variáveis sexo e idade dos informantes não influenciaram no comportamento geoprosódico desses falares mapeados. Além disso, os informantes femininos do ensino médio (Macapá) e fundamental e superior (Belém) não realizaram o contorno ascendente nas interrogativas totais na frase **kwk** (O bisavô gosta do bisavô). Esse desvio no padrão entoacional feminino precisa de uma investigação futura.

Diante dos resultados, retomam-se as hipóteses levantadas na introdução. Uma delas, está relacionada à formação sócio-histórica de Belém e Macapá. Provavelmente, a colonização portuguesa tenha influenciado os falares dessas localidades, próximas geograficamente. Nesse aspecto, os resultados do estudo de Cruz, Seara e Moutinho (2015) que constataram similaridade prosódica de origem açoriana, sinalizam que é possível o contexto histórico influenciar na prosódia de uma comunidade linguística. Confirma-se também a hipótese de haver similaridade prosódica na região nuclear da sentença, pois os resultados de Belém e Macapá apresentaram as mesmas tendências no que concerne à configuração geral das curvas de F0. Mesmo sendo possível observar alguma diferença

no nível do contorno de F0 para o informante do ensino médio nas duas variedades, ambas as cidades têm contornos prosódicos semelhantes. Portanto, os achados das autoras corroboram com este estudo.

Ainda com relação à região nuclear, as análises revelaram que o padrão entoacional de Belém e Macapá está em consonância apenas nas declarativas com o do Rio de Janeiro, descrito por Moraes (1993). O comportamento entoacional das declarativas neutras apresentou pretônicas ascendentes e tônicas descendentes. Com relação às interrogativas totais, os dados rejeitaram a hipótese do movimento circunflexo proposto pelo pesquisador nas duas variedades. Contudo, isso já era esperado por se tratarem de sentenças oxítonas, o que sinaliza que outras pautas acentuais devem ser analisadas. Outrossim, a maioria dos informantes femininos na frase **kwk**, das duas variedades não perfizeram o contorno ascendente na tônica final, o que não corroborou com o padrão encontrado pelo autor. É possível que isso tenha ocorrido por algum artefato metodológico (amostra pequena), o que será melhor investigado.

Quanto à hipótese sobre a relevância do parâmetro acústico da F0 investigado nas duas cidades, de acordo com as análises acústicas aqui empreendidas, este mostrou-se relevante para evidenciar semelhanças prosódicas entre as variedades, uma vez que a pauta acentual oxítona apresentou um contorno entoacional semelhante nas sílabas acentuadas para cada variedade amazônica investigada, o que potencializou com o propósito deste estudo.

De modo geral, esses resultados contribuirão com a pesquisa de Lemos (em andamento) que busca descrever um padrão prosódico caracterizador dessas variedades faladas na Amazônia Oriental, assim como o padrão descrito pelas pesquisas aqui referenciadas na descrição do português do Brasil.

REFERÊNCIAS

AMPER-POR. *Variação Linguística*. Disponível em <www.varialing.eu>. Acesso em 10 de Mai. de 2020.

BARANOWSKI, Maciej. *Sociophonetics*. In: BAYLEY, Richard.; CAMERON, Richard.; LUCAS, Ceil. (eds.). *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press, 2013. p. 403-424.

BOERSMA, Paul; WEENINK, David. *Praat, a system for doing phonetics by computer*, version 5.1.10. Institute of Phonetic sciences of the University of Amsterdam, 2011.

BRITO, Camila. *Acento versus entoação no português falado em Belém (PA): análise prosódica*

dos dados de Belém do Projeto AMPER-Norte. 2014.102f. (Trabalho de Conclusão de Curso). Instituto de Letras, Universidade Federal do Pará, Belém, 2014.

CHAMBOULEYRON, Rafael. *Povoamento, Ocupação e Agricultura na Amazônia Colonial (1640-1706)*. Belém: Ed. Açai/Programa de Pós-graduação em História Social da Amazônia (UFPA) - Centro de Memória da Amazônia (UFPA), 2010.

CUNHA, Cláudia. *Entoação Regional no Português do Brasil*. Tese de Doutorado em Língua Portuguesa, Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2000.

CRUZ, Regina.; BORGES, Benedita.; FERREIRA, Jany.; RILLIARD, Alebert; FONTEL, Emanuel. *Análise acústica dos correlatos prosódicos do acento lexical do português brasileiro*. Joss Journal of Speech Science, v. 5, p. 159-175, 2017.

CRUZ, Regina; SEARA, Isabel; MOUTINHO, Lurdes. “*Intonation of Yes/No question in Portuguese: analysis of the relationship between intonation and lexical stress in the AMPER-POR corpus*”. in: Nebot, Adrián Cabedo (ed.). *Perspectivas actuales en el análisis fónico del habla. Tradición y avances en la fonética experimental*, Valência: Universidade de Valência, pp. 35-44 (ISBN: 978-84-606-8969-0 / ISSN: 2174-7245), 2015.

DESHAIES-LAFONTAINE, Denise. *A Socio-Phonetic Study of a Quebec French Community: Trois-Rivieres*. Tese (Doutorado em Filosofia) – Universidade de Londres: Londres, 1974.

FURLAN, Antônio. *Influência açoriana no português do Brasil em Santa Catarina*. Florianópolis, UFSC, 1989.

LABOV, William. *The Social Stratification of English in New York City*. Washington. D.C: Center for Applied Linguistics, 1966a.

LEMOS, Rosinele. *Descrição geoprosódica do português falado nos centros urbanos da Amazônia Oriental: uma comparação interdialetoal entre Belém, Macapá e São Luís*. Tese (Doutorado em Linguística) – Universidade Federal do Pará, (em andamento), Belém.

LEMOS, Rosinele; CRUZ, Regina. *The dialectal prosodic variation in brazilian portuguese spoken in Baião city (Pa): an intonational analysis of declarative and yes/no interrogative clauses from amper-por corpus*. *Dialectologia. Special Issue*, v. 1, p. 95-108, 2016.

LEMOS, Rosinele; BRITO, Camila. *Análise acústica do português falado na Amazônia: uma*

comparação geoprosódica entre os falares de Belém (PA), Borba (AM) e Macapá (AP). Comunicação apresentada no V Congresso Internacional de Dialectologia e Sociolinguística (CIDS), Universidade Federal da Bahia, 11-14 de setembro 2018.

MORAES, João Antônio. *A Entoação Modal Brasileira: Fonética e Fonologia. Caderno de Estudos Linguísticos*: Campinas, v. 25, p. 101-111, 1993.

NUNES, Vanessa. *Análises Entonacionais de Sentenças Declarativas e Interrogativas Totais nos Falares Florianopolitano e Lageano*. 2015. 178f. Dissertação (Mestrado em Linguística) - Universidade Federal de Santa Catarina, Florianópolis, 2011.

R Development Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL <https://www.R-project.org>.

RILLIARD, Albert. *script lance_batch_easyalign_V3.praat* versão 5.1.10. Limsi, 2018.

RODRIGUES, Aryon. *As Línguas Gerais Sul- Americanas*. *Papia*. n. 4:2, 1996. p. 6-18.

ROMANO, Antonio. *Interface Matlab*. Grenoble, 1999.

SILVA, Joelma. *Caracterização prosódica dos falares brasileiros: as orações interrogativas totais*. 2011, 132f. Dissertação (Mestrado em Linguística) – Faculdade de Letras/UFRJ, Rio de Janeiro, 2011.

SILVESTRE, Aline. *A entoação regional dos enunciados assertivos nos falares das capitais brasileiras*. 2012. 115f. Dissertação (Mestrado em Linguística) – Faculdade de Letras/UFRJ, Rio de Janeiro, 2012.

THOMAS, Erik. *Sociophonetics. an introduction*. New York: Palgrave Macmillan, 2011.

TESTING CHILDREN'S KNOWLEDGE OF GENERIC NULL PRONOUNS¹

TESTANDO O CONHECIMENTO DE CRIANÇAS SOBRE PRONOMES GENÉRICOS NULO

*Karina Bertolino*²

ABSTRACT

In this paper, I discuss an experiment conducted with children acquiring Brazilian Portuguese (BP) as their native language. The experiment was designed to test if Brazilian children understand that the null subject in impersonal structures has the generic reading in BP instead of the referential one. The experiment consisted of a Truth-Value Judgement Task (TVJT). The results show that children as young as 4 years of age understand the null subject in impersonals as generic. Based on a study showing that a 2-year-old child acquiring Estonian already produces null impersonals (TORN-LEESIK; VIJA, 2012), it is possible that children acquiring BP correctly assign the generic reading to a null pronoun in impersonal constructions before the age of 4. I propose that this knowledge could be tested in children younger than 4 using the Intermodal Preferential-Looking (IPL) paradigm (GOLINKOFF *et al.*, 1987; NAIGLES; TOVAR, 2012), a method more suitable than the TVJT to test children that are very young.

Keywords: generic null pronoun, impersonal structures, language acquisition

RESUMO

Neste artigo, discuto um experimento conduzido com crianças adquirindo o português brasileiro (PB) como língua materna. O objetivo do experimento era testar se crianças brasileiras entendem que o sujeito nulo em estruturas impessoais tem uma leitura genérica no PB, em vez de referencial. O experimento consistiu em uma Tarefa de Julgamento de Valor de Verdade (TJVV). Os resultados mostram que crianças com 4 anos de idade já interpretam o pronome nulo em impessoais como genérico. Baseado em um estudo que mostra que uma criança de dois anos adquirindo o estoniano já produz sujeitos nulos impessoais, é possível supor que crianças adquirindo o PB atribuam corretamente a leitura genérica ao pronome nulo de impessoais antes dos 4 anos. Proponho que esse conhecimento possa ser testado em crianças antes dessa idade usando o Paradigma do Olhar Preferencial, um método mais adequado do que a TJVV para testar crianças muito novas.

Palavras-chave: pronome nulo genérico, estruturas impessoais, aquisição da linguagem

1 I would like to express my gratitude to Diane Lillo-Martin and Letitia Naigles for their comments and corrections which improved the paper in a significant way. Thanks also to two anonymous reviewers for *Linguística*, who provided very helpful comments and suggestions. All remaining errors are my own.

2 University of Connecticut, Linguistics Department. E-mail: karina.bertolino@uconn.edu

Introduction

In this paper, I report an experiment using the Truth-Value Judgement Task (TVJT) designed to test if children acquiring Brazilian Portuguese (BP) between the ages of 4 and 7 understand that the null subject in impersonal structures has the generic reading. After discussing the experiment, I outline the design of another experiment, which could be used to test children younger than 4 on the same property discussed here. The experiment I propose uses the Intermodal Preferential-Looking (IPL) paradigm as a method.

There are several types of pro-drop languages. *Radical pro-drop languages*, such as Chinese, Korean and Thai, are characterized by lacking subject-verb agreement. They allow dropping of any argument in the structure, subject or object, in fairly unrestricted circumstances. BP is considered to be a *partial null-subject language*: the null subject is possible under specific and restricted conditions. Partial null-subject languages fall between *consistent null-subject languages*, such as Italian, Spanish, European Portuguese (EP), which allow referential null subjects in any context, and languages that do not allow null-subjects, such as English, German and French. According to Holmberg et al. (2009), for BP, Finnish and Marathi (all partial null-subject languages) the conditions that allow for non-expletive null subjects include:

- (1) a. when the subject is a generic pronoun corresponding to English 'one' (or 'you');
- b. when the subject is controlled by an argument in a higher clause.

In the experiment to be reported, I tested children's knowledge of (1a), using impersonal structures such as (2a).

- (2) a. Nessa escola *e* não pode escovar os dentes depois de comer.

In.this school *e* not can brush:INF the teeth after of eat:INF

'In this school one cannot brush one's teeth after eating.'

Whereas in consistent null-subject languages only the referential reading of the third person pronoun (he/she) is possible in (2a), in partial null-subject languages such as BP, the sentence in (2a) can only be understood as generic: in the place specified by the fronted adverb 'in this school', *one* cannot brush one's teeth after eating. In European Portuguese, a consistent null-subject language, sentence (2a) only allows for the referential reading of the null subject. In order to express the generic reading of the null subject, EP uses a *se* pronoun:

- (2) b. Nessa escola não *se* pode escovar os dentes depois de comer.

In.this school not SE can brush:INF the teeth after of eat:INF

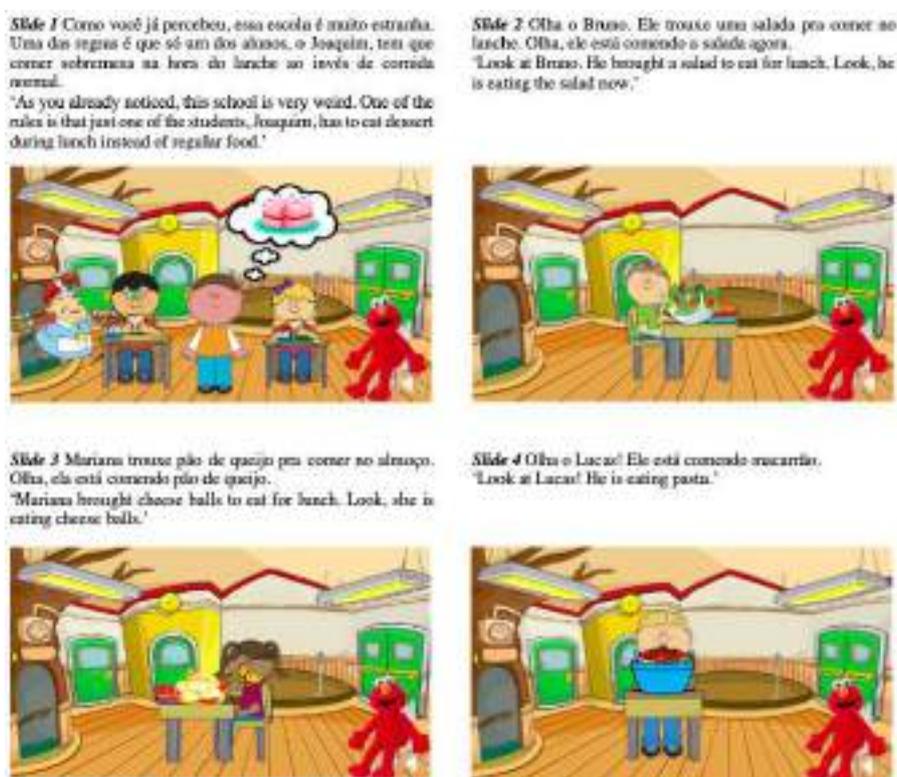
'In this school one cannot brush one's teeth after eating.'

I designed the experiment described below in order to see if Brazilian children know the reading of the null pronoun in a sentence like (2a) is generic. The prediction was *if children knew that BP is not a consistent null-subject language, they should reject the referential reading of the null pronoun in a sentence like (2a)*. As will be discussed in more detail in section 5, the experiment does not have the intention to answer whether Brazilian children know they are acquiring a partial null-subject language. It only intends to show if they know that the null subject in sentences like (2a) is generic: this is so because radical pro-drop languages behave in the same way as partial null-subject languages in allowing only the generic reading of the null pronoun in a sentence like (2a).

1. Methodology and Materials

The experiment employed a Truth-Value Judgement Task (TVJT) methodology (CRAIN; MCKEE, 1985). In this task, the child has to provide a binary judgment: whether the statement made by the puppet is true or false. In our case, children heard a sequence of narratives in BP about a school and judged, at the end of each story, if the character Elmo said something that happened in the story (true) or not (false). PowerPoint animations were used and the task took about 15 minutes to be completed. As mentioned before, the objective of this experiment was to see if children acquiring BP understand that the null subject in an impersonal structure has to be generic instead of referential. If they understand that BP is *not* a consistent null-subject language like EP, the null subject would be judged invariantly as generic. See below an example of one of the stories (Figure 1):

Figure 1: TVJT experiment



Slide 5 Primeiro, o Joaquim colocou um sanduíche na lancheira dele, mas ele lembrou da regra de que ele tinha de trazer sobremesa para o almoço. Então ele trouxe um brigadeiro. Olha, ele está comendo um brigadeiro.
'At first, Joaquim put a sandwich in his lunchbox, but then he remembered the rule that he had to bring dessert for lunch, so he brought a brigadeiro instead. Look, he is eating a brigadeiro.'



Slide 6 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele presta atenção ou não.
'Now Elmo is going to tell us a part of the story. Let's see if he paid attention or not.'



Slide 7 Nessa escola é tem que PRO comer doce na hora do almoço.
In this school has that eat:INF dessert at time of the lunch
'In this school one has to eat dessert for lunch.'



Source: images assembled by the author³

In the story, Joaquim is the possible referential subject. Notice that if the child understands the null subject as referential, (s)he would say that the sentence is *true*, because Joaquim, in fact, had to eat dessert for lunch. If the child understands the null subject as generic, (s)he should say that the test sentence presented in *Slide 7* is *false*. The generic reading of the null subject indicates that it is a property of those that are subject to the rules of the school that they should eat dessert for lunch⁴. However, in the story, this rule was applied to a single character, Joaquim.

The experiment was composed of six test sentences (Table 1), two training items and two fillers, presented randomly between the test sentences. Test sentences were given in two different conditions,

³ Original sources: Jim Henson's Elmo puppet available at <https://www.nbcnews.com/datetime/video/elmo-joins-today-for-tell-a-joke-day-and-matt-lauer-breaks-him-up-1025796675794>; <https://www.amazon.com/Gund-028399758968-25-Elmo/dp/B000EA6YQU>; children, food and objects available at <https://www.mycutegraphics.com/graphics/graphics.html>; waitress, available at <https://filacoahuila.com/10725-waiter-waitress-royalty-free-clip-art-customer-demand-cliparts-aaabee/>; background, available at <https://www.deviantart.com/gamerrainbowshy> (accessed on June 2018)

⁴ If we observe the story that pairs with the test sentence in *Slide 7*, the reading in which the reference of the null subject could be constructed as the specific students from the school (*i.e.* Bruno, Mariana and Lucas) would be pragmatically felicitous. This reading of the null subject is the so-called existential reading and it would be best translated as *they* instead of *one*. However, semantically, the test sentences used in the experiment can be classified as characterizing sentences which have a generic reading, not an existential one. One of the properties of characterizing sentences is that they express a *rule* or a general fact (KRIFKA *et al.* 1995). To illustrate how the test sentences used in the experiment are characterizing sentences, consider, for example, the test sentence presented in *Slide 7*. The sentence expresses a *rule* that holds in the school (students have to eat dessert for lunch). It states that it is a property of those that are subject to the rules of the school that they have to eat dessert for lunch (this happens to be false, given the fact that only Joaquim should eat dessert for lunch).

as shown in Table 1. The different conditions were the type of modal presented in the sentence (*tem que* 'have to' and *não pode* 'cannot'). For half of the test sentences, the expected answer was *true* because the generic reading was *true* in the context; for the other half, the expected answer was *false* because only the referential reading was possible (and this reading is ungrammatical in BP).⁵ All the test sentences were null impersonals with a fronted adverb and a deontic modal, as we can see in Table 1. This structure was used because a survey conducted in Bertolino (2017) showed null impersonals following this pattern were highly accepted by native speakers of BP, but the same did not happen with analogous structures without the fronted adverb and/or the deontic modal⁶.

Table 1: Test sentences

CONDITION	TEST SENTENCES	EXPECTED ANSWER
Tem que ('have to')	Nessa escola <i>e</i> tem que PRO trazer animal de estimação. In.this school <i>e</i> has that bring:INF pets 'In this school one has to bring pets.'	T(true)
Tem que	Nessa escola <i>e</i> tem que PRO comer doce na hora do almoço. In.this school <i>e</i> has that eat:INF dessert in.the time.of.the lunch 'In this school one has to eat dessert for lunch.'	F(false)
Tem que	Nessa escola <i>e</i> tem que PRO brincar dentro da sala de aula. In.this school <i>e</i> has that play:INF inside.of.the room of class 'In this school one has to play in the classroom.'	F
Não pode ('cannot')	Nessa escola <i>e</i> não pode PRO estudar de manhã. In.this school <i>e</i> not can study of morning 'In this school one cannot study in the morning.'	F
Não pode	Nessa escola <i>e</i> não pode PRO escovar os dentes depois de comer. In.this school <i>e</i> not can brush:INF the:PL teeth after of eat 'In this school one cannot brush one's teeth after eating.'	T
Não pode	Nessa escola <i>e</i> não pode PRO entregar a lição de casa. In.this school <i>e</i> not can turn.in:INF the lesson of home 'In this school one cannot turn in the homework.'	T

Source: author

Table 2: Training Items and Fillers

⁵ An anonymous reviewer pointed out that the two conditions combined with the expected answer, which also could be considered a condition, lead to a 2x2 design with four conditions. According to the reviewer, the experiment should have minimally twelve experimental sentences. The experiment reported here only had a total of ten sentences (six test sentences, two training items and two filler items). The number of the sentences used in the experiment were limited by children's attention span. When the pilot was conducted, it was noticed that few children could focus for more than 15 minutes. In the future, though, the experiment could be conducted with more sentences and the same child could be tested on different days.

⁶ The explanation for why null impersonals with a fronted adverb and a deontic modal are more acceptable than their counterparts lacking these elements goes beyond the scope of this paper. But it can be said that the fronted adverb is satisfying the requirement to have [Spec, TP] filled (CHOMSKY, 1981, 1982). The null impersonal pronoun is unable to satisfy this requirement, if we adopt the approach that it is a ϕ P pronoun which stays in [Spec, ν P] (HOLMBERG, 2005). The deontic modal, on the other hand, functions as an overt marker of genericity in the sense that its presence helps us establish that the statement that follows it has a deontic modal base: it describes a world in which a rule holds (see DAHL, 1975; HEIM, 1982 and PAPAFRAGOU, 1996 on the link between genericity and modality).

TRAINING ITEMS	EXPECTED ANSWER
O Joaquim não disse a palavra dinossauro. The Joaquim not say:PAST the word dinosaur 'Joaquim did not say the word dinosaur.'	T(true)
O Joaquim coloriu o carrinho de amarelo. The Joaquim paint:PAST the little.car of yellow 'Joaquim painted the car yellow.'	F(false)
FILLERS	EXPECTED ANSWER
O Joaquim levou uma banana nanica pra escola. The Joaquim take:PAST a banana dwarf to school 'Joaquim brought a dwarf banana to school.'	F
O Joaquim foi de Homem Aranha pra escola. The Joaquim go:PAST of man spider to school 'Joaquim went to school wearing a Spiderman costume.'	T

Source: author

Notice that the two deontic modals used in the test sentences were *tem que* ('have to') and *não pode* ('cannot'). Some valid criticism could be made regarding this choice. First of all, the modal with the negation (*não pode*) should have been avoided since the negation might introduce processing difficulties for the child. Also, it could be argued that as *não pode* is preceded by a negator and *tem que* does not have it, one cannot say that the test sentences are testing equally the same thing (*i.e.* the test sentences do not have two modals, but a modal (*tem que*) and a modal preceded by negation (*não pode*)).

There is a reason why the negation was added to the modal *poder*, though. The modal *tem que* very clearly denotes obligation and imposition (COMPARINI, 2008), while *poder* by itself is very often associated with the epistemic reading, but when preceded by negation, *poder* can be used as a "strong restrictor", a modal that imposes a prohibition (or a rule) and it is, therefore, clearly deontic. Another modal that has a deontic reading is *dever* (must). Nevertheless, this modal can have an epistemic reading in some contexts (see (3) below, in which the modal expresses "possibility"):

(3) ... e nessa hora... é que eu percebi que o que tinha... era uma coisa assim de de de::... uma mistura de sentimento que **devia** ter uma ligação com ciúme... porque eu nunca fui ciumenta... (COMPARINI (2008: 41))

... and at this moment... it was when I noticed what was happening... it was something like... a mix of feelings that **must** have an association with jealousy... because I was never jealous...

In sum, I chose to use *poder* with the negator because (i) it would be problematic to use *poder* by itself, without the negation, since the modal would probably be interpreted as epistemic by the participants; (ii) the other modal that could be used as deontic, namely *dever*, is ambiguous between the epistemic and the deontic reading.⁷

3. Participants

I interviewed forty-two monolingual children acquiring BP as their native language. They ranged in age from 4;0 to 7;10 (mean age = 5;8). The children were recruited in a variety of places: “Museu da Imaginação” (museum), children’s bookstores, activity centers (“Grapali Brinquedoteca”, “Terra do Nunca Brinquedoteca” and “Play Space”)⁸. All children were attending pre-school or elementary school. For analysis purposes, children were divided into four groups by age (each group had 10 children): 4;0 to 4;11 years of age, 5;0 to 5;11 years of age, 6;0 to 6;11 years of age and 7;0 to 7;11 years of age. Fifteen adults were tested as well.⁹

4. Results

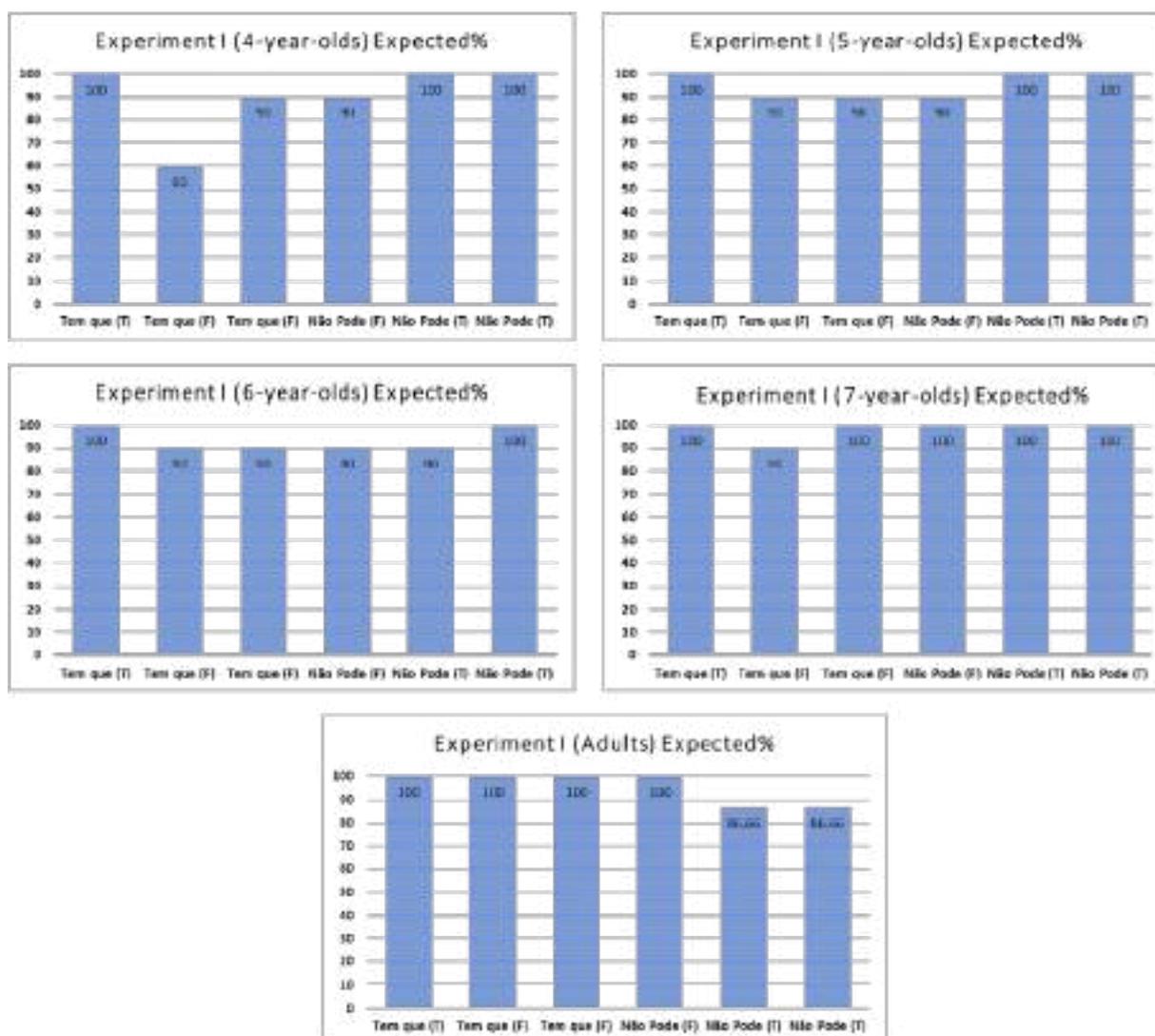
The total number of children interviewed, including the ones that had to be excluded from the analysis, is forty-two. The number of children included in the analysis was forty. One child was excluded because he was too old (8-year-old). Another one (4-year-old) was excluded because he did not want to answer the questions. Children who answered the training items wrongly were corrected until they demonstrated they understood the experiment. Filler items were answered as expected by the children who were included. The results for each age group are presented in the graphs below. Each graph provides the percentage of responses that matched the expected answers for each test-item (in the order they were presented in Table 1, above):

7 Lunguinho (2014) studied the spontaneous speech of two children between 1;02 and 4;11 years of age. He found that the modal *dever* in the children’s grammar only expresses epistemic modality. This is another reason why it would have been problematic to use *dever* in the test sentences as a deontic modal.

8 All children were recruited in the city of São Paulo, Southeast Region of Brazil.

9 Socioeconomic status was not a variable taken into account in this study. That is, no information about the socioeconomic background of children was collected. As the presence of generic null subjects in BP is taken to be the result of a parametric option (precisely, a negative setting of the D in T and ϕ -dependent parameters, as Holmberg (2010) proposes), the prediction is that this grammatical property is not subject to sociolinguistic variation. This prediction needs to be investigated in the future, though.

Graph 1: Percentage of expected answers by each age group (N = 10, for each child group bar chart; N = 15, for the adult bar chart)



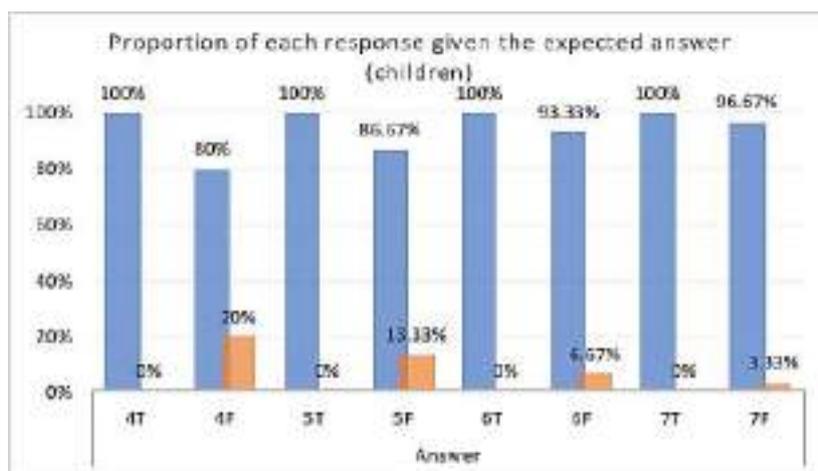
Source: author

As the graphs suggest, children behaved adult-like (supplying more than 85% of expected answers for each item), with the exception of 4-year-olds on the second item with *tem que*. Notice, though, that this happened only for one item, and that the group performance was adult-like otherwise. It could be argued that the reason the rate of expected answers was only 60% for this particular test sentence was because it was the first ‘false’ test sentence which was presented to the children. They had to exclude, for the first time in the test, the referential reading of the null subject. For 4-year-olds, false test sentences could be the hardest items in the whole experiment and these sentences could have caused them some confusion in their first judgement¹⁰. In fact, the statistical analysis shows an

¹⁰ As an anonymous reviewer observed, the high proportion of incorrect answers for the first false item of the experiment could be avoided by (a) randomizing the test items or (b) having multiple test lists, in which the items would be presented in a different order across participants.

effect of the answer (whether it was true or false); children were more likely to give the wrong answer when the expected answer was false (Wilcoxon test: $Z = 2.78$, $p < 0.01$), as can be visualized in the following bar chart (Graph 2). The y-axis quantifies the proportion of true and false answers that were given by the children. The x-axis contains the ages (4-, 5-, 6- and 7-year-olds) and an indication of the expected answer (F(alse) or T(rue)). As we can see, mistakes were *only* made when the expected answer was F (mistakes are indicated by orange, while correct responses are indicated by blue) for all age groups.

Graph 2: Bar chart showing an effect of the answer type (children)



Source: author

A non-parametric Friedman test of differences among repeated measures was conducted to see if there is a difference between the children's age groups and between conditions. Children were not compared to adults, since the level of accuracy of 7- and 6-year-olds was already adult-like, with a percentage of unexpected answers of only 3% for the second group and 1.6% for the first group. The results of the Kruskal-Wallis show there is no significant difference between any age group (Chi-square = 0.734, $p = 0.8651$). In other words, in general, all the groups of children provided expected answers and there is no significant difference between them. No difference between the conditions *tem que* and *não pode* was found either (Friedman Test: Chi-square = 3.00, $p = 0.083$).

5. Discussion

Readers might be concerned children performed well in the experiment not because they have the knowledge that in BP the null subject of impersonals should be generic, but because the referential subject was not salient enough. In our stories the character "Joaquim" was always the possible referential subject. If it was the case this character was not salient enough in the stories, children would have no choice but to choose the generic reading: the one that would always be the

expected answer. Therefore, a failure in the methodology would explain why children performed well in the experiment.

Nevertheless, it doesn't seem to be the case that "Joaquim", the possible referential subject, was not salient enough in the stories. In all the stories, Joaquim was the last character to appear and be mentioned before the slide in which Elmo is about to say the test sentence. That was the case even for the fillers and training items (in these items, Joaquim was mentioned in the sentence itself, as we can see in Table 2). Besides that, Joaquim was salient because he was either the only character that would be the exception to a general rule, or the only character to which a rule would apply. Going back to Table 1, for the impersonal sentence translated into English as, "In this school one has to bring pets", Joaquim is the only one in the school that does not bring a pet. For the sentence, "In this school one has to eat dessert for lunch", Joaquim is the only student that has to eat dessert for lunch, while all the other students eat regular food. In the story, for the sentence, "In this school one has to play in the classroom", while all students play outside, Joaquim is required to play in the classroom. For "In this school one cannot study in the morning" all students are required to study in the morning, while Joaquim has to take a nap during the daytime. In the story, for the sentence, "In this school one cannot brush one's teeth after eating", the students receive a weird recommendation from a dentist that they shouldn't brush their teeth after eating, but Joaquim, who has big teeth, is the only student that has to brush his teeth after eating¹¹. Finally, in "In this school one cannot turn in homework", while all the children are not required to turn in their homework assignments, Joaquim is the only one who is required to do that. In sum, Joaquim is salient enough in the context: not only does he appear and is mentioned in all the slides that precede the test sentence, but he is also the most noticeable character in the stories due to his deviance to the general rules or to the fact that the rules only apply to him, sometimes.¹²

Not being aware of any other confounding factor that could explain why children performed well in this experiment, it can be concluded that children as young as 4 years old reject the referential reading of null subjects in impersonal sentences in BP, suggesting that they already know that they

11 The rules presented in the stories from the experiment were explained to the child to be absurd and inappropriate if applied to the real world before I started each test. The effect was that the stories would sound amusing to the children and they did not seem bored while watching them. Sometimes children would request me to play the stories over again to them.

12 An anonymous reviewer observed Joaquim is no longer the most salient discursive topic when the test sentences are uttered by Elmo, as they are preceded by the referential expression *uma parte da história* 'a part of the story' (*agora o Elmo vai dizer uma parte da história...* 'now Elmo is going to tell us a part of the story'). This topic shift could make it difficult for participants to recover Joaquim as a possible referent for the null pronoun. It is a question, though, whether this topic shift really happens, since differently than 'Joaquim', 'a part of the story' is [-animate] and cannot serve as a referent for the null subject in the test sentences which always describe rules applied to human beings.

are *not* acquiring a consistent null-subject language.

Now I turn to the discussion of whether or not we can conclude that 4-year-olds already know that they are acquiring a partial null-subject language. This conclusion cannot be extracted from this experiment, since radical pro-drop languages (languages without overt verbal agreement) allow only the generic reading of the null subject in sentences equivalent to the ones in the experiment reported here. In the following sentence in Chinese (4), a radical pro-drop language, the subject is null and it can only have the generic reading, the referential reading being impossible¹³ (see also HOLMBERG; PHIMSAWAT, 2005, for Thai):

- (4) Zai zhe-ge xuexiao, e chi fan yihou e bu neng shua ya.
At this-CL school, e eat meal after e not can brush teeth
'In this school one cannot brush one's teeth after eating.'

It can be drawn from the experiment that children as young as 4 years old already know that the null subject in a structure like (2a), repeated here in (5), has the generic reading. Therefore, Brazilian children *know they are not acquiring a consistent null-subject language*, since in consistent null-subject languages such as EP, the null subject in (5) necessarily has the referential reading.

- (5) Nessa escola e não pode escovar os dentes depois de comer.
In.this school e not can brush:INF the teeth after of eat:INF
'In this school one cannot brush one's teeth after eating.'

6 Future research

6.1 Early knowledge of null impersonals: the case of Estonian

In this section I propose a new experiment to test the same knowledge in younger children. Before discussing the new experiment, I would like to explain why it is worthwhile to test the knowledge that children younger than 4 have regarding null impersonals. The motivation for this new experiment is the study by Torn-Leesik and Vija (2012). The authors show that a child acquiring Estonian (a partial-null subject language like BP) uses null impersonals with the generic reading quite early, at the age of 2. This fact opens the possibility that children acquiring BP also demonstrate knowledge of impersonal pronouns before the age of 4.¹⁴

13 I am indebted to Shengyun Gu and Margaret Chui Yi Lee for the Chinese data.

14 A longitudinal study on null impersonal constructions in BP also needs to be conducted in the future. To my knowledge, there is no study specifically addressing the acquisition of null impersonal construction in BP, although there are longitudinal studies conducted on the acquisition of null subjects in BP in general (SIMÕES 1999; MAGALHÃES, 2006).

Torn-Leesik and Vija (2012) present a longitudinal study on the acquisition of Estonian impersonals in the speech of one child: Andreas, age 1;7 to 3;1 (CHILDES database)¹⁵. Estonian is reported to be a partial null-subject language, at least in its spoken form (HOLMBERG, 2016: 366). Estonian impersonals can be formed either with transitive or intransitive verbs, with unaccusatives and modals (TORN, 2002; TORN-LEESIK, 2007, 2009). Notice in the examples below in (6) that Estonian impersonal sentences have no overt subject, which makes this study interesting for our purposes, since the impersonal structures that we are studying in BP also have a null subject.

In Estonian, impersonal structures are marked in the following way: in the present tense, the suffix *-(d/t)akse* on the verb (6a) indicates that the voice in the construction is impersonal. In the past tense, the suffix *-d/ti* marks impersonality on the verb (6b). In the present perfect, the impersonal is formed by the third person present form of the auxiliary *olema* ('to be') and the *-tud* participle marker on the main verb (6c). The past perfect is formed in a similar way to the present perfect, the only difference being that the auxiliary *olema* is in the third person past form (6d)¹⁶:

- (6) a. Loetakse raamatuid.
 read:IMP:PRES books:PART
 'One reads books.'
- b. Loeti raamatuid.
 read:IMP:PAST books:PART
 'One read books.'
- c. On loetud raamatuid.
 be:PRES:3 read:PASS:PTC books:PART
 'One has read books.'
- d. Oli loetud raamatuid.
 be:PAST:3 read:PASS:PTC books:PART
 'One had read books.'

(TORN-LEESIK; VIJA, 2012: 252)

15 CHILDES (Child Language Data Exchange System) <http://childes.psy.cmu.edu/>

16 The following is a list of abbreviations which appear in (6) and (7): IMP impersonal, ILL illative, PART partitive, PASS passive, PRES present, PTC particle, 3 third person.

The use of null impersonals in Estonian with the generic interpretation was found very early in the speech of Andreas. Consistent adult-like use of the construction starts at the age of 2;3 and it progresses until the last age analyzed (3;1), although Andreas sometimes has slight problems in using the correct verb stem, as we can see from the example below:

- (7) siis näsutakse [*] [= näsutatakse] ja siis pannakse prügikasti.
 then chew:IMP:PRES and then put:IMP:PRES bin:ILL
 ‘then one chews it and then puts it in the bin.’ (Andreas, 2;8)
 (TORN-LEESIK; VIJA, 2012, p.262)

It might be the case that null impersonals were acquired early by Andreas due to particularities in the language: according to Viitso (1998), null impersonals are considered to be basic constructions in Estonian. It does not seem, though, that the same is true for BP. According to the adult data reported by Assis (2017), the generic null pronoun is being replaced by the generic pronoun *você*, which suggests that null impersonals are not a “basic construction” in BP and therefore they may not be as salient in the input as null impersonals in Estonian are.

Testing null impersonals in 3- and 2-year-olds acquiring BP is important because it would show whether (i) the early acquisition of null impersonals in Estonian is a language-particular phenomenon or (ii) it is a more general pattern that can be seen across different languages allowing generic null pronouns. Also, since the study in Estonian had only one participant, it would be important to test the acquisition of impersonals with a group of children.

6.2 An IPL experiment

The TVJT is not an appropriate method to test children younger than 3;0 (BLUME; LUST, 2017). Even when this age restriction is adopted by studies, younger children exhibit more acceptance of false scenarios and ‘yes bias’ (EISELE; LUST, 1996). As argued by Blume and Lust (2017: 148), the TVJT requires complex cognitive computation by participants. The child participating in a TVJT experiment must perform the following cognitive tasks: (i) (s)he must listen to a story and interpret it; (ii) (s)he must hear the test sentence and relate it to a relevant part of the story just told; (iii) (s)he must judge whether the test sentence is true or false. Relating the test sentence to a relevant part of the story (*i.e.* step (ii)) can be challenging to young children: we can imagine the scenario in which the child says that the test sentence is *false* even when it is *true*, because it does not match a certain part of the story the child perceived as relevant.¹⁷ Also, when asking the child to judge whether the test

17 Take as an example the test sentence “In this school one has to bring pets”. The sentence was true, because the

sentence is true or not we need to rely on the child's sense of "truth" (*i.e.* the correspondence between a situation and what is said).

There is another type of TVJT in which no story is told to the child: pictures are presented, and the child has to judge whether a sentence is true or not given the pictures. This version of the TVJT is simpler than the dynamic version of the task with a story, in the sense that the child does not need to listen to a story and interpret it. Children still must relate the test sentence to a relevant part of the picture depending on the complexity of the image (*i.e.* step (ii)). In this version of the TVJT, we still have to rely on the child's sense of "truth" (*i.e.* step (iii)).

As I intend to test 3- and 2-year-olds, there is the need to use a method more suitable to young children. The Intermodal Preferential Looking (IPL) paradigm¹⁸ is a method that does not demand any verbal response, acting-out or pointing by children, which makes it suitable to assess language comprehension of very young children (as young as 9 months) (NAIGLES; TOVAR, 2012). Besides having the advantage that the child is not asked to give any explicit response to the experimenter, the IPL paradigm does not require the child to perform any of the complex cognitive tasks just described. First of all, as we will see below (Figure 2), the child does not have to listen to a story and interpret it (*i.e.* step (i)) during an IPL experiment. The IPL paradigm also does not run into the same issues of relevance of information as the TVJT (*i.e.* step (ii)), as the child does not have to choose a part of a story or image to judge whether it is true or not. Most importantly, the IPL does not rely on the child's notion of "truth" and in her/his judgement (*i.e.* step (iii)): the child does not have to judge whether a sentence is true or false, (s)he just has to look at the clip that matches the test sentence. For these reasons, the IPL paradigm is arguably a simpler task than the TVJT and thus more appropriate to test children who are below the age of four.

The first version of the IPL paradigm was created by Golinkoff *et al.* (1987) and the version of the IPL I present here is based on Naigles and Tovar (2012). In the IPL paradigm, children watch side-by-side video clips while a camera records their eye-movements. The clips appear in series starting with a non-directing audio (for control purposes), and then with a test audio that should match only one of the videos. A screen with a blinking red light appears between stimuli to hold children's attention. If the child wants to sit on the parent's lap, the parent cannot have auditory or visual access to the stimuli. In order to not influence the child's choice, the parent listens to music on a mp3 player while watching the video.

After the data is collected, children's eye movements are coded offline by coders who are blind to the conditions. The basic prediction is that children who have the linguistic knowledge about the

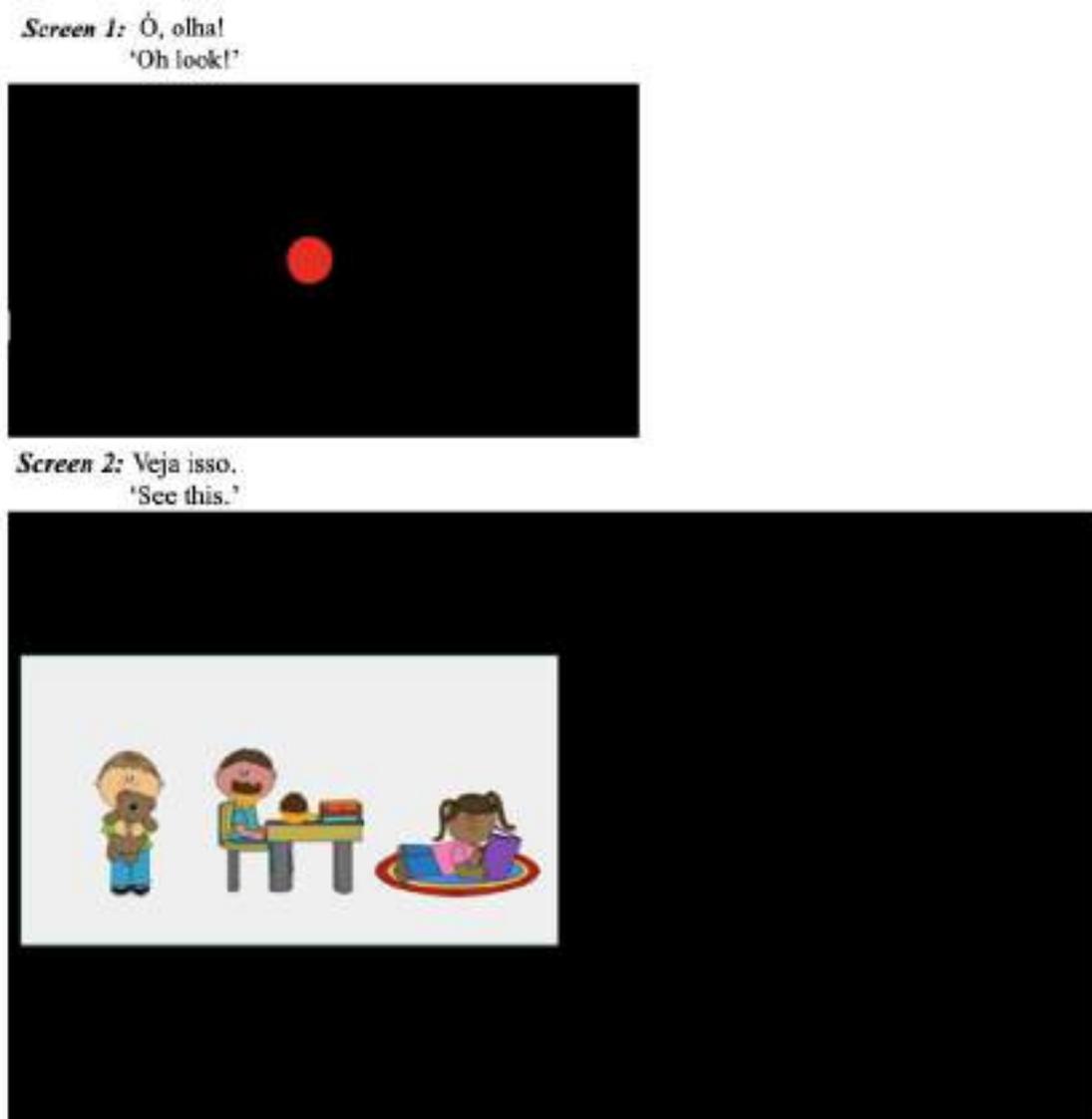
students in the school had to bring pets after the first day of school. However, a child who stays fixated on the events that occurred on the *first day of school* might say that the sentence is false instead of true, because no one had to bring pets on the first day of class.

18 As an anonymous reviewer pointed out, the IPL is a method that has been used in Brazil by two research groups: LAPAL (PUC-Rio) and NEALP (UFJF).

construction being tested will look more quickly and longer at the video that matches the test audio.

The objective of this experiment is the same as in the one previously reported: to see if children acquiring BP understand that the null subject in an impersonal structure has to be generic instead of referential, with the difference that with the IPL experiment we will be able to test much younger children. If children understand BP is *not* a consistent null-subject language, they should look more quickly and longer at the clip compatible with the generic reading of the pronoun (showing a collective of people performing the action described in the impersonal sentence) than at the clip compatible with the referential reading of the pronoun (showing a single individual performing the action described in the impersonal sentence). Ideally, we should use test sentences very similar to the ones used in the TVJT, in order to obtain comparable results in both tasks. Figure 2 shows an example on how the IPL experiment would work:

Figure 2: IPL Experiment



Screen 3: Ó, olha agora!
'Oh, look now!'



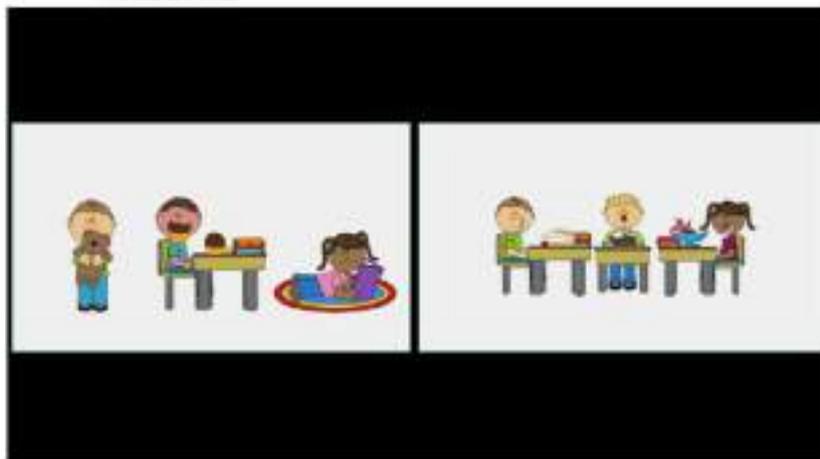
Screen 4: Veja aqui.
'See here.'



Screen 5: Ei, olha aqui!
'Hey, look here!'



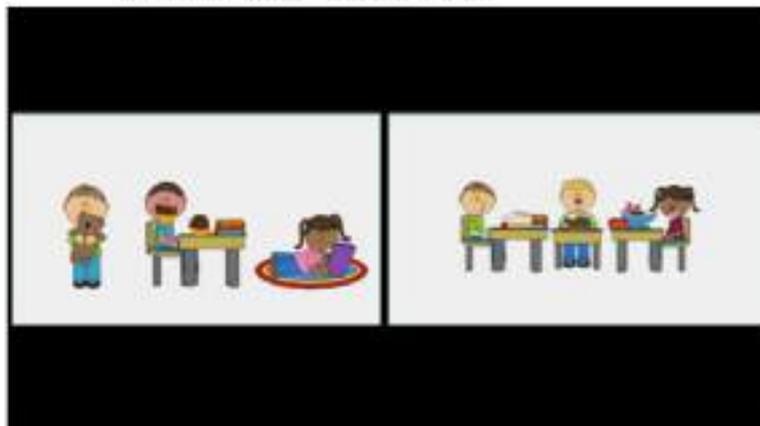
Screen 6: A gente pode ver os dois!
'We can see both!'



Screen 7: Nessa escola e tem que PRO comer doce.
 In.this school has that eat:INF dessert
 'In this school one has to eat dessert.'



Screen 8: Nessa escola e tem que PRO comer doce.
 In.this school has that eat:INF dessert
 'In this school one has to eat dessert.'



Source: author

Source: author, on children and objects available at:
<https://www.mycutegraphics.com/graphics/graphics.html>

The first four screens are paired with a familiarization audio with the clips alternating between right and left. Besides having the function to familiarize the child with the experiment, the first four screens also teach the child how the experiment will work, making her/him look to the direction in which the clip appears. *Screen 6* corresponds to the baseline trial: here the test stimulus is presented for the first time, but it is not paired with the test-sentence: it is paired with a non-directing audio instead. The function of the baseline trial is to show whether the child has any visual preference for either stimulus. When analyzing the data, we can tell with the baseline trial whether the results are due to a visual bias or linguistic knowledge. In *Screen 7*, the child hears the test sentence for the first time without the visual stimulus (with a red blinking light). In *Screen 8*, the test sentence is paired with the test stimulus. If the child knows that the referential reading of the pronoun is blocked in BP, the child will look more quickly and longer at the clip that shows “students in general” eating dessert (generic reading) than at the clip showing just one student eating dessert (referential reading).

In contrast to the TVJT task presented previously, the images prepared for the IPL task intend to be simple, not distracting for 2- and 3-year-olds from the main point: choose whether or not the null subject can have the referential reading. This is why the images are presented in a white background. Another important point is that, in order to avoid the preference for one particular screen over another, we keep the number of characters constant in both screens and randomize the target side.

Concluding Remarks

In the present paper, I discussed results of an experiment conducted with children acquiring BP as their native language, from the age of 4;0 to 7;0. The experiment used the TVJT as a method. My objective was to see (i) if children knew that the null subject in an impersonal construction has the generic reading in BP and also (ii) if children would rule out the referential reading of the null pronoun, not available for native speaker adults of BP in the constructions tested.

The results of this experiment show that children acquiring BP as young as 4 years of age correctly assign the generic reading to the null subject in impersonal sentences and rule out the referential reading of the null subject. With these results, I concluded that 4-year-olds already know BP is *not* a consistent null subject language. I then discussed if we could conclude that all 4-year-olds already know that BP is a partial null-subject language. As we expect children would exhibit the same performance if they thought they were acquiring a radical pro-drop language (*i.e.* rejecting the referential reading of the null subject in the constructions tested and allowing the generic reading of the null subject), this conclusion cannot be extracted from this experiment. It can only be concluded that 4-year-olds acquiring BP already know they are *not* acquiring a consistent null-subject language such as EP.¹⁹

Based on the fact that the age of acquisition of null impersonal constructions is reported to be 2 years old in Estonian (TORN-LEESIK; VIJA, 2012), I proposed an experiment using the Intermodal Preferential-Looking (IPL) paradigm to test 2- and 3-year-olds acquiring BP. This experiment has the potential to show whether (i) null generic pronouns in Estonian are acquired early because impersonal

¹⁹ As an anonymous reviewer observed, a more complete study on the acquisition of impersonal structures in BP should take into account the existential reading of the null pronoun in the language, not only the generic one. As Carvalho (2019) discusses, sentences such as *Naquela loja vendeu café por muito tempo* ('Someone sold coffee in that store for quite some time') have an existential reading in BP, that is, they have the interpretation that a non-specified person sold coffee in a certain store. Existential impersonals with a null subject are possible in BP, a partial null-subject language, but not in consistent null-subject languages such as EP. Therefore, it would be relevant to conduct a study seeing how children acquiring BP interpret existential impersonals. In the study presented here, impersonals could not be interpreted as existential, due to the presence of the deontic modal and the law-like background present in all the test sentences and their stories (see note 4).

is regarded as a basic construction in Estonian or (ii) null generic pronouns are acquired early across different languages.

If it can be demonstrated that children acquiring BP have a very early knowledge of null impersonals, as happens in Estonian, we can ask how children attain this knowledge. How could we explain the early knowledge of generic pronouns? How do children learn that a certain pronoun can express an abstract property such as genericity? If 2-year-olds already have this sophisticated knowledge, could it not be the case that all children start out with a grammar which allows generic null subjects and blocks referential ones whenever generic null subjects are possible?²⁰ If children start out with a grammar that allows generic pronouns, that would implicate that children acquiring consistent null-subject languages incorrectly assign the generic reading to the null subject at some point (as if they were acquiring a radical or partial null-subject language) and later shift to a consistent null-subject grammar in which the null subject has the referential reading. How about children acquiring non-null-subject languages such as English or Swedish? Do they, at some point in their development, allow generic null subjects? In order to test the hypothesis that children start out with a grammar that allows generic null subjects it would be interesting to conduct the IPL experiment proposed in section 6 with children acquiring consistent and non-null-subject languages, in addition to conducting it with children acquiring partial null-subject languages such as BP.

REFERENCES

ASSIS, Jacqueline de Sousa Borges de. *De se a (vo)cê: o percurso da indeterminação do Português Brasileiro*. Tese (Doutorado em Linguística) - Programa de Pós-Graduação em Estudos Linguísticos, Universidade Federal de Uberlândia, Uberlândia, 2017.

BERTOLINO, Karina. *Modals and Impersonal Null-Subjects in Brazilian Portuguese*. Dissertação (Mestrado em Linguística), Connecticut, University of Connecticut, 2017.

BLUME, María; LUST, Barbara. *Research methods in language acquisition: Principles, procedures, and practices*. Berlin: Walter de Gruyter, 2017.

CARVALHO, Janayna. Incorporated subjects in existential impersonal sentences in Brazilian Portuguese. In: Herbeck, Peter; Pöll, Bernhard; Wolfsfruber, Anne. (org.), *Semantic and syntactic aspects of impersonality*. Helmut Buske Verlag, 2019.

²⁰ A similar question is asked by Gelman *et al.* (2008) regarding generic NPs (e.g. *birds lay eggs*): generic expressions are a challenging problem for the child, since they are abstract (*i.e.* one cannot point to kinds, only to instances of a kind). Gelman *et al.* (2008) propose that generic NPs are the default form and non-generic NPs need to be learned by the child.

CHOMSKY, Noam. Lectures on Government and Binding. Dordrecht: Foris, 1981.

_____. *Some concepts and consequences of the theory of Government and Binding*. Chicago: MIT Press, 1982.

CRAIN, Stephen; MCKEEN, Cecile. The acquisition of the structural restrictions on anaphora. In: Berman, Stephen; Choe, Jae-Woong; and McDonough, Joyce. (org.), *Proceedings of NELS 16*. Amherst, MA: GSLA. p. 94-110, 1985.

COMPARINI, Ana Maria Paulino. *A natureza multissignificativa dos verbos modais: uma análise da língua falada no Brasil. Diálogos Pertinentes*. Revista Científica de Letras. São Paulo: UNIFAN, v. 4, n. 4, p. 31-46, 2008. Disponível em <https://doi.org/10.26843/dp.v4i1.220>. Acesso em 30 de dez. 2019.

DAHL, Östen. On generics. In KEENAN, Edward. (org.), *Formal semantics of natural language*, 99-112. London and New York: Cambridge University Press, 1975.

EISELE, Julie, and LUST, Barbara. *Knowledge about Pronouns: A Developmental Study Using a Truth-Value Judgment Task*. Child Development, v. 67, n. 6, p. 3086–3100, 1996. Disponível em <www.jstor.org/stable/1131768>. Acesso em 30 de dez. 2019.

GELMAN, Susan; GOETZ, Peggy; SARNECKA, Barbara; FLUKES, Jonathan. *Generic Language in Parent-Child Conversations*. Language Learning and Development, v. 4, n. 1, 1-31, 2008

GOLINKOFF, Roberta; HIRSH-PASEK, Kathy; CAULEY, Kathleen; GORDON, Laura. *The eyes have it: Lexical and syntactic comprehension in a new paradigm*. Journal of Child Language, v. 14, n. 1, 23-45, 1987.

HEIM, Irene. *The semantics of definite and indefinite noun phrases*. Tese (Doutorado em Linguística), University of Massachusetts, Amherst, 1982.

HOLMBERG, Anders. *Is there a little pro? Evidence from Finnish*. Linguistic Inquiry, v. 36, n. 4, p. 533-564, 2005. Disponível em <https://www.jstor.org/stable/4179340?seq=1>. Acesso em 30 de dez. 2019.

HOLMBERG, Anders. Null subject parameters. In BIBERAUER, Theresa; HOLMBERG, Anders; ROBERTS, Ian; SHEEHAN, Michelle (org.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press, 2010.

HOLMBERG, Anders. Linguistic Typology. In ROBERTS, Ian. (org.), *The Oxford Handbook of Universal Grammar*, 354–76. Oxford University Press, 2016.

HOLMBERG, Anders; NAYUDU, Aarti; SHEEHAN, Michelle. *Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi*. *Studia Linguistica*. v. 63, n. 1, p. 59-97, 2009. Disponível em <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9582.2008.01154.x>. Acesso em 30 de dez. 2019.

HOLMBERG, Anders; PHIMSAWAT, Ony. *Generic pronouns and phi-features: evidence from Thai*. *Newcastle and Northumbria Working Papers in Linguistics (Second Asian and European Linguistics Conference Proceedings)*, v. 21, n.1, 2005. Disponível em <https://pdfs.semanticscholar.org/f69f/ec7cc5e6b836b581f7af28b7bda43c90aea2.pdf>. Acesso em 30 de dez. 2019.

KRIFKA, Manfred; PELLETIER, Francis Jeffrey; CARLSON, Gregory; ter MEULEN, Alice; GODEHARD, Link; CHERCHIA, Gennaro. Genericity: an introduction. In: CARLSON, Gregory and PELLETIER, Francis Jeffrey (org.), *The generic book*, p. 1-124. Chicago and London: University of Chicago Press, 1995.

LUNGUINHO, Marcus Vinicius. *On the acquisition of root and epistemic modals in Brazilian Portuguese*. *ReVEL*, special issue 8, 2014.

MACWHINNEY, Brian. *The CHILDES Project: Tools for analyzing talk*. Lawrence Erlbaum Associates, Inc. 1991.

MAGALHÃES, Telma Moreira Vianna. *O sistema pronominal sujeito e objeto na aquisição do português europeu e do português brasileiro*. Tese (Doutorado em Linguística) - Programa de Pós-Graduação em Linguística, Campinas, Universidade Estadual de Campinas, 2006.

NAIGLES, Letitia; TOVAR, Andrea. *Portable Intermodal Preferential Looking (IPL): Investigating Language Comprehension in Typically Developing Toddlers and Young Children with Autism*. In: *Journal of Visualized Experiments*, v. 70, e4331, 2012. Disponível em <https://www.jove.com/video/4331/portable-intermodal-preferential-looking-ipl-investigating-language>. Acesso em 30 de dez. 2019.

PAPAFRAGOU, Anna. *On generics*. *University College London Working Papers in Linguistics*, v. 8, n.1, p. 165-198, 1996.

SIMÕES, Luciene. *Sujeito Nulo na Aquisição do Português do Brasil*. *Cadernos de Estudos*

Linguísticos, Campinas, v. 36, n. 1, p. 105-130, jan./jun. 1999. Disponível em <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8637107>. Acesso em 31 de dez. 2019.

TORN-LEESIK, Reeli; VIJA, Maigi. *Acquisition of the Impersonal Voice by an Estonian Child*. Journal of Baltic Studies. v. 43, n. 2, 251-271, 2012. Disponível em <https://doi.org/10.1080/01629778.2012.674799>. Acesso em 30 de dez. 2019.

TORN-LEESIK, Reeli. *Voice and Modal Verbs in Estonian*. Linguistica Uralica, v. 18, n. 3, p. 173–86, 2007.

TORN, Reeli. *The Status of the Passive in English and Estonian*. In: Hendriks, Henriëte. (ed.) RCEAL Working Papers in English and Applied Linguistics. Cambridge: Research Centre for English and Applied Linguistics. v. 7, n.1, p. 81-106, 2002

_____. *The Voice System of Estonian*. Sprachtypologie und Universalienforschung, v. 62, n.1 p. 2–90, 2009

VIITSO, Tiit-Rein. Fennic. In: ABONDOLO, Daniel (org.) *The Uralic Languages*, p. 96-114. London: Routledge, 1998

ESTRUTURAS DE POSSE INALIENÁVEL ADJETIVADA: APONTAMENTOS INICIAIS SOBRE A SUA AQUISIÇÃO EM PORTUGUÊS BRASILEIRO

ADJECTIVAL INALIENABLE POSSESSION STRUCTURES: INITIAL NOTES ON ITS ACQUISITION IN BRAZILIAN PORTUGUESE

Fernanda Mendes¹

RESUMO

A presente pesquisa tem como objetivo aprofundar a análise sintático-semântica de estruturas possessivas inalienáveis apresentada em Mendes (2017, 2015), observando dados de aquisição do português brasileiro comparados a dados da gramática-alvo. Para tanto, examina-se o seu comportamento quando ocorrem em construções predicativas e atributivas (BARON; HERSLUND, 2001) modificadas por um adjetivo, já que parece haver, conforme aponta o trabalho de Kayne (1975) para o francês, restrições para a veiculação da leitura inalienável de acordo com a presença e/ou tipo de adjetivo que modifica o nome possuído nessas construções com possuidor externo. Assume-se nesta pesquisa, estendendo o que Guéron (1985) e Vergnaud e Zubizarreta (1992) afirmam também para o francês, que as restrições propostas por Kayne (1975) para essas construções relacionam-se com a estrutura temática em português brasileiro. No que concerne ao processo de aquisição da linguagem, enquanto as restrições não são adquiridas, hipotetiza-se que a interpretação inalienável seja superestendida, tornando-se restrita às condições licenciadas pela gramática adulta à medida em que a criança avança em idade. Para verificar as hipóteses adotadas nesta pesquisa, foi desenvolvido, portanto, um estudo de caráter experimental com três grupos, sendo dois de informantes infantis (crianças mais jovens e mais velhas) e um de informantes adultos (controle). Os resultados, que, por um lado, confirmam a hipótese aventada para o percurso da aquisição dessas estruturas, e, por outro, não se apresentam como conclusivos em relação à gramática adulta do português brasileiro, indicam que, embora não categórico, o presente estudo traz importantes apontamentos acerca da questão.

Palavras-chave: posse inalienável, adjetivos, aquisição da linguagem, teoria gerativa, português brasileiro.

ABSTRACT

This research aims at advancing the syntactic-semantic analysis of inalienable possession structures presented in Mendes (2017, 2015), by observing Brazilian Portuguese acquisition data compared to its target grammar data. To that end we examined these constructions' behavior when they occur in predicative and attributive structures (BARON; HERSLUND, 2001) modified by an adjective, because there seem to be, following Kayne's work (1975) on French, some constraints that block inalienable interpretation due to the presence and/or the kind of adjective modifying the possessed noun when there is an external possessor in this kind of construction. We assume, by extending Guéron (1985) and Vergnaud & Zubizarreta (1992)'s claims about French, that Kayne's constraints (1975) concerning adjectival possessive structures are related to their own thematic structure in Brazilian Portuguese. As regards the language acquisition process, although these constraints are not acquired, we hypothesize that inalienable interpretation is overextended, becoming restricted to the conditions licensed by the adult grammar. In order to test the working hypothesis, we advanced an experiment with three groups, two with child informants (younger and older children) and one with adult informants (control group). The results borne out the working hypothesis on the acquisition of these structures, but are not conclusive in relation to the adult Brazilian Portuguese grammar, indicating that, although not being categorical, the present case-study brings important insights on the issue.

Keywords: inalienable possession, adjectives, language acquisition, generative theory, Brazilian Portuguese.

¹ Pesquisadora do Programa de Pós-Doutorado da Universidade Federal de São Carlos e bolsista do PNPd-CAPES. E-mail: fernanda.mds@gmail.com.

Introdução

Construções possessivas, segundo Baron e Herslund (2001, p.3), podem ser divididas em dois grandes tipos, a saber, construções de posse predicativa e construções de posse atributiva.

Nas construções de posse predicativa, a ligação possessiva entre duas entidades – possuidor e possuído – é realizada através de verbos como *possuir* e *ter*, como ilustram os exemplos em (1) abaixo. Por esse motivo, é também chamada de *posse estabelecida*, conforme Seiler (1983², p.62 *apud* BARON e HERSLUND, 2001, p.13).

- (1) a. O João *tem* um livro.
b. A Maria *possui* um carro.

Já nas construções de posse atributiva, a ligação possessiva entre possuidor e possuído, ao invés de ser realizada por um verbo possessivo, é pressuposta por outro elemento que indique posse, tal qual um pronome possessivo, como ilustram os exemplos em (2). Por isso, é chamada também de *posse inerente*, de acordo com Seiler (1983, p.62 *apud* BARON e HERSLUND, 2001, p.13).

- (2) a. O João vendeu o livro *dele*.
b. A Maria dirigiu o carro *dela*.

Assim, conforme Mithun (2001, p.285), enquanto a posse predicativa abrange sintagmas verbais (ou sentenciais), a posse atributiva abarca sintagmas nominais.

Segundo uma restrição proposta por Kayne (1975) para dados do francês, construções de posse predicativa com possuidor externo ao sintagma nominal possuído exigem a ocorrência de um adjetivo descritivo para que a leitura inalienável³ possa ser veiculada, como em (3a) abaixo – contrastando com (3b), em que não há ocorrência de qualquer adjetivo. Por outro lado, construções de posse atributiva, com possuidor externo ao sintagma nominal possuído proíbem a ocorrência de tais adjetivos para a veiculação da mesma leitura, como em (4a) abaixo – contrastando com (4b), em que se trata de uma construção de posse atributiva com possuidor interno ao sintagma nominal possuído.

- (3) a. Elle a les cheveux *bruns*.
 ela-NOM.3SG ter-PRS.3SG os.DEF.PL cabelos.PL castanhos.PL
 ‘Ela tem os cabelos *castanhos*.’

2 SEILER, H. Possession as an operational dimension of language. Tübingen: Gunter Narr, 1983.

3 Quando o elemento possuidor e o elemento possuído estabelecem, entre si, uma relação intransferível, como em (i) e (ii) abaixo, em que o elemento possuído *cabeça* faz parte do corpo do elemento possuidor O João e não pode ser transferido a nenhum outro possuidor; e o elemento possuído *mãe* estabelece uma relação de parentesco com o elemento possuidor A Maria que não pode ser transferida a outrem.

- (i) O João levantou a *cabeça*.
(ii) A Maria abraçou a *mãe*.

- b. #Elle a les cheveux.
 ela-NOM.3SG ter-PRS.3SG os.DEF.PL cabelos.PL
 ‘Ela tem os cabelos.’
- (4) a. #Il adore les cheveux *blondes*.
 ele-NOM.3SG adorar-PRS.3SG os.DEF.PL cabelos.PL loiros.PL
 ‘Ele adora os cabelos *loiros*.’
- b. Il adore *ses* cheveux *blondes*.
 ele-NOM.3SG adorar-PRS.3SG seus-GEN.PL cabelos.PL loiros.PL
 ‘Ele adora *seus* cabelos *loiros*.’

Entretanto, como apontado por Mendes (2017, 2015), ao menos em relação ao comportamento de construções de posse inalienável não-adjetivas, a gramática do português brasileiro (doravante PB) não apresenta exatamente as mesmas restrições apresentadas pela gramática do francês.

Assim, a presente pesquisa visa examinar se essa restrição acerca da ocorrência de adjetivos em construções de posse inalienável também se aplica aos dados do PB adulto assim como é aplicada aos dados do francês, como ilustram os dados em (6) e (7) abaixo, bem como se propõe a investigar de que forma se dá a aquisição de tais estruturas possessivas em PB.

- (5) a. #O João *tem* os dedos.
 b. O João *tem* os dedos *lindos*.
 c. O João *tem* os dedos *quebrados*.
- (6) a. O João levantou os dedos.
 b. ?#O João levantou os dedos *lindos*.
 c. O João levantou os dedos *quebrados*.

Os exemplos em (5) e (6) acima tratam de construções com possuidor externo ao sintagma nominal possuído, sendo os exemplos em (5) de construções de posse predicativa, enquanto os exemplos em (6) de construções de posse atributiva. Dessa forma, parece ser possível observar alguns dos efeitos da restrição sobre a ocorrência de adjetivos nessas estruturas possessivas em relação à possibilidade de veiculação da leitura inalienável em PB, uma vez que os dados em (5a-6a) são construções sem nenhum tipo de adjetivo, em que a leitura inalienável, de fato, parece ser bloqueada. Já os dados em (5b-6b) são construções com adjetivos descritivos, em que a leitura inalienável deveria ser bloqueada, no entanto, esse bloqueio interpretativo não parece estar tão assegurado em PB, quanto parece estar em francês. Por fim, os dados em (5c-6c) são construções com adjetivos restritivos, em que a leitura inalienável, de fato, parece ser veiculada.

Partindo dos pressupostos da Gramática Gerativa e assumindo a Hipótese Inatista (CHOMSKY, 1981, 1986), este estudo apoia-se, especialmente, em trabalhos que discutem estruturas de posse inalienável e sua aquisição, tais como Mendes (2017, 2015, 2010a, 2010b, 2010c), Floripi e Nunes (2009), Munn *et al.* (2006), Schaeffer e Mathewson (2005), Pérez-Leroux *et al.* (2004, 2002a,b), Schaeffer (2002). Além desses, esta pesquisa também tem como base estudos que investigam a restrição acerca de adjetivos dentro dessas estruturas possessivas, tais como Vergnaud e Zubizarreta (1992), Authier (1988), Guéron (1985) e Kayne (1975), para explicar as diferentes interpretações disponíveis para as estruturas inalienáveis adjetivas em PB.

Hipotetiza-se, neste trabalho, que, estruturas possessivas inalienáveis adjetivas em construções de posse atributiva com o possuidor externo ao sintagma nominal possuído, como em (6b-c), apresentariam, na gramática infantil do PB, uma preferência pela leitura inalienável mesmo quando estiver presente um adjetivo descritivo no sintagma nominal possuído, tal qual o fariam na presença de um adjetivo restritivo. Além disso, prevê-se que estruturas possessivas inalienáveis adjetivas em construções de posse predicativa com o possuidor externo ao sintagma nominal possuído, como (5b-c), também apresentariam, na gramática infantil do PB, uma preferência pela leitura inalienável mesmo sem qualquer adjetivo presente na estrutura do sintagma nominal possuído, tal como o fariam na presença de um adjetivo. Para tanto, tomam-se os achados de Mendes (2017, 2015) acerca da aquisição de estruturas possessivas inalienáveis em PB, combinados aos resultados provenientes de Mendes (2010b) sobre a aquisição de estruturas inalienáveis adjetivas com verbos possessivos/copulares em PB.

Essa combinação entre a presença e/ou os tipos de adjetivos permitidos em cada tipo de construção possessiva e a possibilidade de veiculação das leituras alienável e inalienável seria, portanto, restringida à medida em que a criança converge à gramática adulta.

2. Estruturas de posse inalienável adjetivada

Examinando ambos os tipos de construções no francês, como em (3) e (4), Kayne (1975) propõe restrições para a veiculação da leitura inalienável em estruturas possessivas nas quais ocorrem adjetivos dentro do sintagma possuído. Isto é, nos casos em que o possuidor e o possuído estabelecem entre si uma relação intransferível (leitura inalienável), tal como no caso da relação parte-todo com nomes de partes do corpo ou da relação de instituída por nomes de parentesco, como em (7a-b) abaixo, parece haver restrições interpretativas que não ocorrem quando essa relação intransferível não se faz presente (leitura alienável), tal como no caso dos exemplos em (7c-d).

- (7) a. O João_i lavou as *mãos* t_i.
 b. A Maria_i abraçou o *pai* t_i / *pro*_j.
 c. O João lavou o *copo*.
 d. A Maria abraçou o *menino*.

Segundo Vergnaud e Zubizarreta (1992), a diferença entre essas duas leituras, inalienável e alienável, se dá, entre outras razões, pela propriedade que os nomes de partes do corpo e de parentesco têm de poder realizar o seu possuidor como uma categoria vazia anafórica ou pronominal coindexada a um possuidor externo ao sintagma possuído. Enquanto nomes que não pertencem a essas classes não apresentam essa possibilidade.

Assim, Kayne (1975), mais especificamente, propõe restrições que bloqueiam a veiculação da interpretação inalienável em estruturas em que haja uma relação intransferível estabelecida entre o possuidor e o possuído, quando o possuidor em questão se encontra em uma posição externa ao sintagma nominal possuído, para duas situações:

- (i) No caso de construções atributivas, em que haja a ocorrência de um adjetivo descritivo modificando o nome possuído, ou;
 (ii) No caso de construções predicativas, em que seja observada a ausência de qualquer adjetivo modificando o nome possuído.

Os exemplos em (4a) e (3b) acima, repetidos, por conveniência, em (8) e (9) abaixo, ilustram, respectivamente, essas duas situações.

Construção atributiva com adjetivo descritivo

- (8) #Il adore les cheveux blondes.
 ele-NOM.3SG adorar-PRS.3SG os.DEF.PL cabelos.PL loiros.PL
 ‘Ele adora os cabelos *loiros*.’

Construção predicativa sem qualquer adjetivo

- (9) #Elle a les cheveux.
 ela-NOM.3SG ter-PRS.3SG os.DEF.PL cabelos.PL
 ‘Ela tem os cabelos.’

Caso o PB se comporte tal qual o francês em relação à possibilidade de veiculação de interpretação inalienável em tais estruturas possessivas com o possuidor externo ao sintagma nominal possuído, seria esperado que a ocorrência de adjetivos descritivos (tal como *lindos*) fosse, por um lado, proibida em construções atributivas, tal como o dado em (10b) abaixo, e, por outro lado, exigida em construções predicativas, tal como o dado em (11b) abaixo, refletindo os julgamentos interpretativos possíveis na gramática-adulta.

Construções atributivas (sem verbo possessivo)

- (10) a. O João levantou os dedos⁴ *Leitura inalienável possível*
 b. ?#O João levantou os dedos lindos *Leitura inalienável degradada(?)*
 c. O João levantou os dedos quebrados *Leitura inalienável possível*

Construções predicativas (com verbo possessivo)

- (11) a. #O João tem os dedos *Leitura inalienável degradada*
 b. O João tem os dedos lindos *Leitura inalienável possível*
 c. O João tem os dedos quebrados *Leitura inalienável possível*

Resumidamente, pode-se observar no Quadro 1 abaixo as distinções entre os tipos de posse, sua possibilidade de coocorrência com adjetivos e as possíveis interpretações possessivas obtidas dessas estruturas.

Quadro 1 - Tipos de posse, adjetivos e interpretação possessiva em PB

TIPO DE POSSE	ADJETIVOS	INTERPRETAÇÃO POSSESSIVA
Predicativa	Descritivos	Inalienável
	Restritivos	Inalienável
	Sem adjetivos	Alienável
Atributiva	Descritivos	Alienável
	Restritivos	Inalienável
	Sem adjetivos	Inalienável

Fonte: elaboração própria

Em relação aos dados de construções atributivas com o possuidor externo ao sintagma nominal possuído, ainda não é possível ter uma posição mais assegurada em relação ao bloqueio da interpretação inalienável em PB, uma vez que isso parece depender também da pluralidade do nome possuído. Entretanto, como se observa em (12) – em que (12a) mostra nomes inalienáveis de partes do corpo naturalmente singulares, (12b-c) mostram nomes inalienáveis de partes do corpo naturalmente plurais, e (12d) mostra nomes de partes do corpo naturalmente massivas –, tem-se a ocorrência de dados como (13), que parecem se assemelhar ao que é afirmado sobre o francês.

- (12) a. #A Maria machucou a cabeça linda.
 b. ?#A Maria levantou as mãos lindas.
 c. ?#A Maria levantou os dedos lindos.
 d. A Maria escovou os cabelos lindos.

- (13) #O João adora os dedos lindos.

4 Faz-se importante destacar, nos exemplos em (10), que, mesmo ocorrendo um verbo tal como *levantar*, que parece favorecer uma plausibilidade derivada de conhecimento de mundo, uma vez que é mais plausível levantar-se os próprios dedos, ainda assim as construções atributivas com adjetivos descritivos (10b) não parecem permitir a leitura inalienável, fortalecendo as hipóteses adotadas nessa pesquisa.

De acordo com o estudo de Vergnaud e Zubizarreta (1992), que também analisa dados do francês, essa restrição se deve à natureza aberta ou fechada da expressão nominal que o adjetivo modifica.

Assim, os sintagmas inalienáveis em construções com o possuidor externo ao sintagma nominal possuído são expressões abertas, como apresentadas acima em (7a-b). Isto é, suas variáveis argumentais não são saturadas no DP, mas em VP ou IP, já que o possuidor é o sujeito da sentença (ou o argumento dativo⁵). Ademais, pelo fato de esses sintagmas serem expressões abertas, não podem funcionar como argumentos.

No caso das construções com possuidor interno ao sintagma nominal possuído, como os exemplos em (14), os sintagmas inalienáveis são expressões fechadas. Ou seja, suas variáveis argumentais são saturadas no DP pelo pronome possessivo. Por este motivo, podem funcionar como argumentos.

- (14) a. O João levantou os dedos *dele*.
 b. O João levantou os dedos lindos *dele*.
 c. O João levantou os dedos quebrados *dele*.

Esses autores ainda afirmam que adjetivos descritivos são predicados que tomam argumentos. Assim, modificam apenas expressões fechadas, já que esses adjetivos baseiam-se na relação argumental e essas são as únicas que funcionam como argumento.

Por outro lado, adjetivos restritivos não são predicados que tomam argumentos. Assim, modificam tanto expressões abertas – nas construções com possuidor externo ao sintagma possuído –, quanto expressões fechadas – nas construções com possuidor interno ao sintagma possuído.

Diferentemente dos adjetivos descritivos, os adjetivos restritivos baseiam-se em composição, isto é, o significado do modificador é composto com o do elemento modificado, criando um novo *denotatum*.

Conforme Crisma (1990⁶, 1993⁷ *apud* PRIM, 2010), adjetivos pré-nominais são interpretados

5 Embora não sejam construções muito usuais em PB, são bastante produtivas em francês e espanhol e podem ser exemplificadas abaixo em (i), (ii) e (iii):

- | | | | | | |
|-------|-------------------------|------|---------------|-----------|-------------------|
| (i) | João se lava as mãos. | (PB) | | | |
| (ii) | Jean se lava as mãos. | | lave | les | mains. (Francês) |
| | Jean-NOM se-DAT.SG | | lavar-PRS.3SG | as.DEF.PL | mãos.PL |
| | ‘João se lava as mãos.’ | | | | |
| (iii) | Juán se lava las manos. | | lava | las | manos. (Espanhol) |
| | Juán-NOM se-DAT.SG | | lavar-PRS.3SG | as-DEF.PL | mãos.PL |
| | ‘João se lava as mãos.’ | | | | |

como descritivos, enquanto adjetivos pós-nominais são, de forma geral, interpretados como restritivos.

Isso explicaria, por um lado, uma maior aceitabilidade da interpretação inalienável quando os adjetivos ocorrem em posição pós-nominal, como em (15b-d), dado que é possível que os adjetivos estejam sendo interpretados como restritivos – o que não feriria a restrição proposta por Kayne (1975).

Por outro lado, na posição pré-nominal, como em (15a'-e'), parece, de fato, haver uma interpretação inalienável mais degradada, além de parecerem dados de estilo literário – sendo, assim, pouco naturais para o PB falado. Isso poderia indicar que o PB se sujeita, ao menos em partes, à restrição de adjetivos em construções atributivas, de forma semelhante ao francês.^{6 7}

(15) a. #A Maria machucou a cabeça grande (#e não a pequena).

a'. #A Maria machucou a grande cabeça.

b. A Maria levantou a mão bonita (e não a feia).

b'. ?#A Maria levantou a bonita mão.

c. A Maria levantou os dedos grossos (e não os finos).

c'. ?#A Maria levantou os grossos dedos.

d. A Maria escovou os cabelos longos (?e não os curtos).

d'. A Maria escovou os longos cabelos.

e. #O João adora os dedos finos (#e não os grossos).

e'. #O João adora os finos dedos.

Isto é, nos exemplos acima, quando em posição pós-nominal, o adjetivo pode ser, além de descritivo, restritivo, e, no último caso, não se sujeitaria à restrição das estruturas inalienáveis adjetivas em construções atributivas. Já, quando em posição pré-nominal, o adjetivo apenas pode ser descritivo, e, portanto, não permitiria a veiculação da leitura inalienável, de acordo com a restrição proposta por Kayne (1975) para construções atributivas.

Em uma pesquisa de caráter experimental, reportada por Mendes (2010b), realizada com adultos falantes nativos de PB, foi aplicado um experimento-piloto com o objetivo de testar a compreensão acerca dessas estruturas. Observou-se que, embora não isento de falhas, o experimento com verbos não-possessivos gerou resultados que confirmaram em parte o que a literatura afirma sobre a interpretação disponível para estas estruturas.

No entanto, nesse experimento não foram analisadas as diferenças concernentes aos diferentes tipos de nomes inalienáveis em relação à sua pluralidade, tampouco foi usado algum verbo tal qual

6 CRISMA, P. *Functional categories inside the noun phrase: a study on the distribution of nominal modifiers*. 1990. Tese. University of Venice.

7 CRISMA, P. *On Adjective Placement in Romance and Germanic Event Nominals*. In: *Rivista di Grammatica Generativa*, 18:61-100, 1993.

adorar. Além disso, também não foram exploradas as diferentes posições que o adjetivo pode ocupar, o que torna a presente pesquisa um avanço, além de um aprofundamento, em relação à análise de tais estruturas.

Em relação aos dados de construções predicativas com o possuidor externo ao sintagma possuído, apresentados em (11) acima – e repetidos em (16) abaixo por conveniência –, observa-se, de fato, um bloqueio para a veiculação da interpretação inalienável quando não há um adjetivo modificando o nome possuído. Isso é ilustrado, especificamente, em (16a).

- (16) a. #O João tem os dedos.
b. O João tem os dedos lindos.
c. O João tem os dedos quebrados.

Diferentemente das construções atributivas, nas construções predicativas, a pluralidade dos nomes possuídos não parece ser um fator relevante para a veiculação da interpretação inalienável, como ilustram os dados em (17) abaixo.

- (17) a. #A Maria tem a cabeça.
b. #A Maria tem a(s) mão(s).
c. #A Maria tem os dedos.
d. #A Maria tem os cabelos.

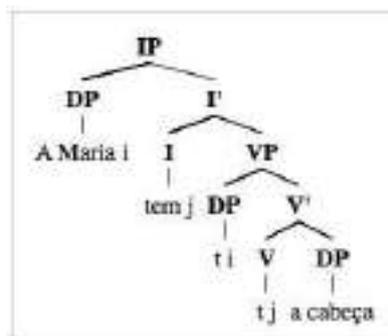
De acordo com Guéron (1985), que também analisa dados do francês, o verbo *ter* apresenta uma estrutura de dupla categorização: (i) como verbo lexical transitivo e (ii) como verbo funcional.

Como verbo lexical transitivo, como nos exemplos em (16a) e em (17), *ter* atribui um papel temático primário a cada um de seus argumentos – externo, *O João*, em (16a), e *A Maria*, em (17), e interno, *os dedos*, em (16a) e *a cabeça/a(s) mão(s)/os dedos/os cabelos*, em (17). No entanto, de acordo com o Critério Temático Revisado pela autora⁸, não é possível formar uma cadeia com dois NPs recebendo papéis temáticos primários e, por isso, a veiculação de leitura inalienável não é permitida, resultando na leitura alienável atestada também em PB. A estrutura arbórea a seguir ilustra essa configuração sintática.

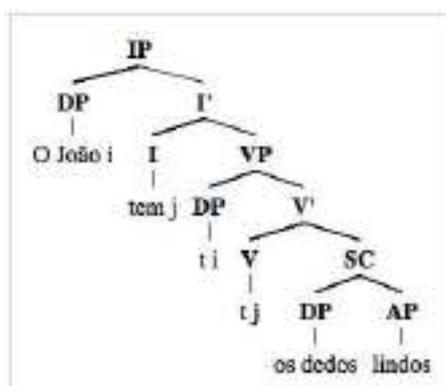
8 Critério temático revisado (GUÉRON, 1985)

- (i) A todo NP com conteúdo lexical é atribuído um e apenas um papel temático.
(ii) Em LF, toda cadeia contém um único argumento e a toda cadeia é atribuído um único papel temático PRIMÁRIO.

Os papéis temáticos primários seriam, segundo a proposta da autora, AGENTE, TEMA, META e FONTE. Os secundários seriam S-META e S-LOCATIVO.



Como verbo funcional, como nos exemplos em (16b-c), *ter* governa uma *small clause* e não atribui papel temático primário à posição de sujeito da sentença, mas um papel temático secundário disponível para ser atribuído a esta posição. Sendo o predicado da *small clause*, ocupado por um adjetivo, responsável por atribuir um papel temático para o sujeito desta mesma *small clause*, ocupado por um nome inalienável. A estrutura arbórea a seguir ilustra essa configuração sintática.



Assim, sendo respeitado o Critério Temático Revisado proposto pela autora, é possível que haja uma cadeia entre os dois NPs – possuidor e possuído –, uma vez que eles não recebem, ambos, papéis temáticos primários, e, conseqüentemente, a veiculação da leitura inalienável é atestada também nos dados do PB.

Em relação ao verbo *ter* em PB, estudos como os de Ribeiro (1996), Avelar e Callou (2007) e Avelar (2009) discutem sobre seu estatuto funcional e, embora discordem em relação à estrutura específica subjacente deste verbo, concordam em relação à ideia de que há, de fato, uma mesma estrutura subjacente para as construções possessivas, copulativas e existenciais em PB, equiparando, por exemplo, o verbo *ter* ao verbo *estar (com)*. Isso poderia se configurar como uma evidência em direção à análise apresentada anteriormente para as estruturas inalienáveis adjetivas em construções predicativas.

A respeito dos adjetivos que ocorrem em construções predicativas, observa-se que a literatura

menciona apenas adjetivos descritivos como necessários à estrutura para que a veiculação da interpretação inalienável seja possível. Entretanto, nota-se que dados contendo adjetivos classificados como restritivos também veiculam a interpretação inalienável.

De forma semelhante ao que Crisma (1990, 1993) afirma acontecer com adjetivos classificados como descritivos em posição pós-nominal, em que eles podem ser interpretados como restritivos, parece acontecer com os adjetivos classificados como restritivos em posição predicativa, uma vez que, nesta posição, podem ser interpretados de forma descritiva e não mais de forma restritiva, como ilustram os exemplos em (18) abaixo.

- (18) a. A Maria tem a mão congelada (#e não a descongelada).
- b. A Maria tem as unhas quebradas (#e não as inteiras).
- c. A Maria tem os cabelos sujos (#e não os limpos).

Muito embora esses mesmos adjetivos não possam ocupar a posição pré-nominal, destinada aos adjetivos classificados como descritivos de fato, como ilustram os dados em (19) abaixo, evidencia-se que esses são adjetivos restritivos que podem veicular uma leitura descritiva.

- (19) a. *A Maria tem a congelada mão.
- b. *A Maria tem as quebradas unhas.
- c. ??*A Maria tem os sujos cabelos.

Os dados em (20) mostram que, em se tratando de uma *small clause*, mesmo que pouco naturais em PB falado, esses adjetivos podem aparecer em posição anterior, porém, não no interior do sintagma possuído.

- (20) a. A Maria tem congelada a mão.
- b. A Maria tem quebradas as unhas.
- c. A Maria tem sujos os cabelos.

Interessantemente, em construções predicativas com o possuidor interno, parece também haver uma diferença interpretativa em relação à posição do pronome possessivo, como mostram os dados em (21). Os exemplos em (21a-b) parecem veicular a leitura alienável e os dados em (21a'-b') parecem veicular a leitura inalienável.

- (21) a. A Maria tem os dedos quebrados dela.
- a'. A Maria tem os dedos dela quebrados.
- b. A Maria tem os dedos lindos dela | os lindos dedos dela.
- b'. A Maria tem os dedos dela lindos.

Quanto aos resultados obtidos por meio do experimento-piloto supracitado aplicado em adultos falantes nativos de PB, observou-se que, com verbos possessivos, as afirmações da literatura sobre a

interpretação disponível para estas estruturas não são completamente confirmadas.

No entanto, por ter se configurado como pesquisa inicial, nesse experimento não foram analisadas as diferentes posições ocupadas pelo adjetivo e as diferentes configurações contendo o possuidor interno. Dessa forma, uma vez mais, torna a presente pesquisa um avanço em relação à análise de tais estruturas.

Assim, estendendo o que Guéron (1985) afirma para os dados do francês, propõe-se, para os dados do PB, que a restrição proposta por Kayne (1975) para estruturas inalienáveis adjetivas esteja relacionada com a estrutura temática dessas construções.

Em estruturas inalienáveis adjetivas com verbos não-possessivos, adjetivos descritivos não são permitidos, pois atribuem outro papel temático ao NP parte do corpo, que, por sua vez, já recebe um papel temático do verbo lexical e, portanto, ferem o Critério Temático.

Assim, o adjetivo descritivo *lindos*, por exemplo, pode ser caracterizado como um adjetivo adnominal predicador de núcleo, de acordo com as pesquisas de Negrão *et al.* (2008) e Müller *et al.* (2002). Ele atribui um papel temático ao NP parte do corpo, que já recebe um papel temático do verbo lexical.

Já o adjetivo restritivo *quebrados*, por exemplo, não atribui um papel temático ao NP parte do corpo, podendo ser caracterizado como um adjetivo predicativo, de acordo com a classificação proposta nas pesquisas das autoras supracitadas. Por este motivo, este adjetivo é permitido nessa construção, da mesma forma que são permitidas as construções sem adjetivos, nas quais o NP parte do corpo recebe apenas um papel temático – aquele atribuído pelo verbo.

Por outro lado, em estruturas inalienáveis com verbos possessivos, adjetivos descritivos são permitidos, pois atribuiriam um papel temático ao NP parte do corpo, que não recebe nenhum papel temático do verbo, que, por sua vez, não se configura como lexical, mas como funcional.

Assim, o adjetivo descritivo *lindos*, por exemplo, diferentemente do contexto anterior, é permitido, pois ele atribui um papel temático ao NP parte do corpo, que não recebe um papel temático do verbo, já que este se comporta como um verbo funcional e seleciona uma *small clause*, conforme Guéron (1985).

O adjetivo restritivo *quebrados*, neste caso, é classificado como adnominal predicador de núcleo, de acordo com a classificação de Negrão *et al.* (2008) e Müller *et al.* (2002), diferentemente de como é classificado no contexto anterior (em que aparece em construções com verbos não-possessivos). Já que, se neste caso também fosse um adjetivo predicativo, não seria permitido nas construções com

verbos possessivos e copulativos, pois não teria um papel temático para atribuir ao NP parte do corpo, que precisa receber um papel temático.

No caso das construções com verbos possessivos e copulativos sem adjetivo, pelo fato de não haver nenhum papel temático sendo atribuído ao NP parte do corpo pelo verbo, que se comporta como funcional, nem pelo adjetivo, que não está presente, perde-se a interpretação inalienável.

A tarefa da criança seria, portanto, restringir o uso dos diferentes tipos de adjetivos, aparentemente livres em PB, de acordo com a estrutura temática das construções em que eles ocorrem, para adequar as interpretações disponíveis na sua gramática à gramática-alvo.

Sabendo-se que esta não é uma tarefa simples, uma vez que depende de restrições sintático-semânticas muito sutis relacionadas ao tipo de adjetivo e ao tipo de construção possessiva, esta pesquisa visa aprofundar os estudos acerca da posse inalienável por meio de um estudo experimental descrito na seção a seguir.

3. Estudo experimental

Tendo em vista a restrição proposta por Kayne (1975) para o francês, a presente pesquisa lança mão de um estudo experimental que toma por hipótese a possibilidade de o PB se comportar tal qual o francês em relação à possibilidade de veiculação de interpretação inalienável em estruturas possessivas predicativas e atributivas contendo adjetivos e um possuidor externo ao sintagma possuído. Dessa forma, seria esperado que a ocorrência de adjetivos descritivos fosse, por um lado, proibida em construções atributivas, e, por outro lado, exigida em construções predicativas. Assim, embora haja uma gama de variações nesse tipo de estrutura possessiva, este trabalho delimitou-se a um recorte inicial, mas não menos importante, que analisa as construções predicativas e atributivas apenas com possuidores externos ao sintagma possuído coocorrendo com nomes inalienáveis de partes do corpo⁹ e, quando presente, o adjetivo descritivo ou restritivo em posição final, a fim de conceder robustez ao teste-piloto já aplicado anteriormente.

3.1. Participantes

Para a coleta dos dados experimentais foi aplicado um teste off-line a um total de 38 participantes¹⁰, que compuseram o *Grupo Experimental (GE)* e o *Grupo Controle (GC)* deste experimento, conforme detalhado na Tabela 1 abaixo.

⁹ A presente pesquisa conta apenas com testes envolvendo nomes inalienáveis de partes do corpo. Os testes envolvendo nomes inalienáveis de parentesco serão realizados em uma pesquisa futura.

¹⁰ Agradeço às instituições que permitiram o desenvolvimento do estudo: CECI Integral/DEdIC/UNICAMP e PRODECAD/DEdIC/UNICAMP.

Tabela 1 - Total de participantes

Informantes	GE1		GE2			GC	Total
	4;0	5;0	6;0	7;0	8;0	Adultos	
Total	8	7	3	4	6	10	38
	15		13				

Fonte: elaboração própria

Os participantes infantis foram divididos em dois grupos, *GE1* e *GE2*. Sendo que o *GE1* foi composto por crianças mais jovens, de quatro e cinco anos de idade, sendo sete meninas e oito meninos, enquanto o *GE2* foi composto por crianças mais velhas, de seis a oito anos de idade, sendo nove meninas e quatro meninos. A escolha dessa faixa etária deve-se às pesquisas anteriormente realizadas sobre o tema, tais como Mendes (2017, 2015, 2010a, 2010b, 2010c), que mostram a conversão da gramática infantil em direção à gramática adulta, isto é, a aquisição das leituras disponíveis às estruturas inalienáveis em PB, por volta do cinco e seis anos de idade.

Já os participantes adultos compuseram o *GC*, tendo entre 18 e 30 anos de idade, que, além de tornar possível a análise da gramática adulta em relação a esses tipos de construções possessivas – uma vez que não há este tipo de estudo para o PB –, também serve como referência de gramática-alvo com a qual se estabelecerá um paralelo tendo em vista os resultados infantis.

A aplicação dos testes foi permitida mediante autorização dada pelo Comitê de Ética em Pesquisa (CEP) – vide o Certificado de Apresentação para Apreciação Ética (CAAE) número 04440918.8.0000.8142, disponível online no site da Plataforma Brasil. Além disso, cada uma das escolas também autorizou a aplicação dos testes e, por fim, os pais de cada uma das crianças assinaram um Termo de Consentimento Livre e Esclarecido autorizando a participação delas nos experimentos, bem como a utilização dos seus dados nesta pesquisa. No caso dos adultos, um Termo de Consentimento Livre e Esclarecido também foi assinado por cada um dos participantes autorizando a aplicação dos testes e a utilização dos seus dados neste trabalho.

3.2. Metodologia

O teste que será descrito a seguir foi baseado na Tarefa de Seleção de Figura (TSF), um método experimental usado para recolhimento de dados de compreensão e caracterizado por Gerken & Shady (1996) como “uma tarefa de compreensão em que ao ouvinte é apresentado um estímulo linguístico (por exemplo, uma palavra ou uma sentença) e é pedido para que ele selecione, de um conjunto de figuras, aquela que corresponde melhor a ele”¹¹ (p. 125). Esse método foi adotado neste trabalho para verificar a interpretação semântica de contrastes morfossintáticos específicos, uma vez que se

11 Do original: “a comprehension task in which the listener is presented with some linguistic stimulus (e.g., a word or a sentence) and asked to select from a set of pictures the one that best corresponds to it” (GERKEN & SHADY, 1996:125).

procurou “inferir a natureza das representações morfosintáticas infantis examinando os tipos de erros de compreensão que elas cometem”¹² (p. 126).

O experimento consistiu em um jogo simples, em que o objetivo era passar pelas “fases” (compostas por sentenças-teste e distratores), mostrando para a experimentadora qual imagem dentre as apresentadas seria a mais adequada em relação à sentença proferida. Assim, o participante conseguia o maior número de estrelas e concorria ao troféu disponível no final da brincadeira. Dessa forma, pôde-se entreter as crianças de maneira efetiva durante o teste e evitar que elas ficassem entediadas.

Para contextualizar os elementos do experimento – personagens, cenas, sentenças-teste e distratores –, a experimentadora contou uma história, apresentando os personagens e o que eles iriam fazer, bem como explicou o jogo e proferiu as sentenças-teste e distratores.

A história versava sobre quatro animais – o coelho, o elefante, o jacaré e o leão – que eram amigos e foram brincar juntos no tanque de areia de um parque. Lá encontraram uma caixa cheia de brinquedos e, ao abrir essa caixa, os animais viram uma série de partes do corpo de brinquedos que eles resolveram usar para se fantasiar.

Assim, enquanto os animais brincavam com o que eles haviam encontrado, o participante e a experimentadora iniciavam o jogo, que é o experimento em si, observando o que iria acontecer com os personagens da história.

No total, o teste exibiu 32 telas aos participantes, sendo que 24 delas continham sentenças-teste, como as apresentadas em (22)-(27) abaixo, e oito delas continham distratores, como os apresentados em (28) abaixo.

Construções predicativas com adjetivo descritivo

- | | |
|--|---|
| (22) a. Me mostra quem tem as orelhas compridas. | Nome inalienável duplo (Npc2) ¹³ |
| b. Me mostra quem tem os dentes brancos. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem tem a juba laranja. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem tem a tromba grande. | Nome inalienável único (NpcU) |

12 Do original: “to infer the nature of children’s morphosyntactic representations by examining the types of comprehension errors that they make” (GERKEN & SHADY, 1996:126).

13 Essas siglas serão utilizadas para a apresentação e análise dos resultados. É importante salientar, aqui, que “Npc” se refere à “nome de parte do corpo”, uma vez que, em um futuro estudo, pretende-se abordar a comparação destes com nomes relacionais (“Nr”).

Construções predicativas com adjetivo restritivo

- | | |
|---|-----------------------------------|
| (23) a. Me mostra quem tem o olho sujo. | Nome inalienável duplo (Npc2) |
| b. Me mostra quem tem o dedo sujo. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem tem o bigode sujo. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem tem o focinho sujo. | Nome inalienável único (NpcU) |

Construções predicativas sem adjetivo

- | | |
|---|-----------------------------------|
| (24) a. Me mostra quem tem as sobrancelhas. | Nome inalienável duplo (Npc2) |
| b. Me mostra quem tem as unhas. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem tem a juba. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem tem a tromba. | Nome inalienável único (NpcU) |

Construções atributivas com adjetivo descritivo

- | | |
|--|-----------------------------------|
| (25) a. Me mostra quem sujou as orelhas compridas. | Nome inalienável duplo (Npc2) |
| b. Me mostra quem sujou os dentes brancos. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem sujou a juba laranja. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem sujou a tromba grande. | Nome inalienável único (NpcU) |

Construções atributivas com adjetivo restritivo

- | | |
|--|-----------------------------------|
| (26) a. Me mostra quem sujou o olho machucado. | Nome inalienável duplo (Npc2) |
| b. Me mostra quem sujou o dedo machucado. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem sujou o bigode cortado. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem sujou o focinho machucado. | Nome inalienável único (NpcU) |

Construções atributivas sem adjetivo

- | | |
|---|-----------------------------------|
| (27) a. Me mostra quem sujou as sobrancelhas. | Nome inalienável duplo (Npc2) |
| b. Me mostra quem sujou as unhas. | Nome inalienável múltiplo (Npc>2) |
| c. Me mostra quem sujou a juba. | Nome inalienável massivo (NpcM) |
| d. Me mostra quem sujou a tromba. | Nome inalienável único (NpcU) |

Distratores

- | | |
|--------------------------------------|--|
| (28) a. Me mostra quem está molhado. | |
| b. Me mostra quem está sujo. | |

Como se pode observar, as sentenças-teste resultam da interação de três variáveis, formando um teste no formato 2x4x3:

- Construção possessiva: *predicativa* ou *atributiva*.
- Unicidade do nome inalienável: *duplo*, *múltiplos*, *massivo* e *único*.
- Tipo de adjetivo: *descritivo*, *restritivo* e *ausente*.

Tanto as sentenças-teste quanto os distratores foram proferidos pela experimentadora tão logo a tela com as imagens correspondentes era apresentada ao participante. Em ambos os casos, esta tela era composta por quatro imagens. No entanto, no primeiro caso, uma das imagens correspondia à leitura alienável, outra à leitura inalienável e as duas restantes a distratores. Já no último caso, apenas uma das imagens correspondem à sentença proferida, enquanto as outras três imagens restantes eram distratores. Os pares de Figuras 1 e 2 ilustram isso



Fonte: elaboração própria

Os participantes escolhiam as imagens que consideravam mais adequadas à sentença proferida por meio de apontamento ou verbalmente. Não foi estipulado um tempo mínimo ou máximo para a resposta, porém todos os participantes levaram um curto período para tomar a sua decisão e, por ser uma tarefa curta e simples, não houve desistências. A escolha foi registrada pela experimentadora em uma folha de papel que continha o questionário do teste impresso. Por fim, as sentenças-teste e distratores foram apresentados em ordem aleatória, bem como as imagens correspondentes às leituras possíveis também foram trocadas de lugar ao longo do experimento, evitando, dessa forma, algum tipo de enviesamento nos resultados obtidos.

4. Resultados e discussão

Levando-se em consideração o design e o público-alvo deste experimento, os resultados advindos da aplicação do teste descrito acima foram analisados de acordo com a *análise livre* das respostas. Isto é, foram consideradas todas as opções de interpretação escolhidas pelo informante – ainda que grande parte dos informantes infantis tenha feito apenas uma escolha interpretativa.

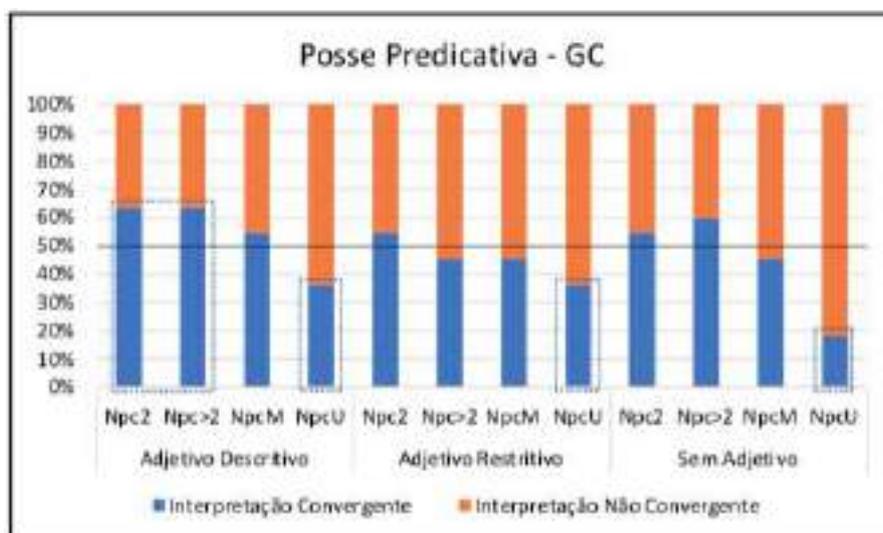
A seguir, segue-se a análise dos resultados, que será apresentada de acordo com o tipo de construção possessiva, sendo exibidos, em primeiro lugar, aquela referente à posse predicativa e, em seguida, aquela que diz respeito à posse atributiva.

4.1. Posse predicativa

Retomando as hipóteses atribuídas à posse predicativa, caso o PB adulto se comportasse tal qual outras línguas românicas em relação a esse tipo de construção possessiva, seria esperado que a leitura inalienável fosse preferencial quando presente, ao menos, o adjetivo descritivo na estrutura. Já no caso de ausência de adjetivos na estrutura, seria prevista a preferência pela leitura alienável.

No entanto, diferentemente do esperado¹⁴ de acordo com a hipótese adotada, como se observa no Gráfico 1 abaixo, a gramática adulta do PB parece exibir, de forma geral, um comportamento *chance* em relação às interpretações disponíveis para esse tipo de construção possessiva. Observa-se uma preferência um pouco mais marcada em duas situações: (i) quando se trata de um nome inalienável contável não-único em construções com adjetivo descritivo, tais como as sentenças em (22a,b,c), e (ii) quando se trata de um nome inalienável único, especialmente em construções sem adjetivo, tal como (24d), e, ainda assim, contrariamente à hipótese aventada.

Gráfico 1¹⁵ - Resultados gerais do Grupo Controle em construções de Posse Predicativa



Fonte: elaboração própria

14 Aqui, a *interpretação esperada* ou *convergente* é tomada como aquela que seria a interpretação escolhida, de acordo com a hipótese adotada. Variando entre alienável ou inalienável, conforme é modificada a construção sintática e as suas condições em relação ao que afirma a teoria.

15 Em todos os gráficos, “C” abrevia interpretação *convergente* e “NC” abrevia a interpretação *não convergente*, sendo essas leituras especificadas em cada um dos gráficos de análise. Além disso, retoma-se que “Npc” se refere a “Nomes de Partes do Corpo”, sendo “Npc2” os nomes de partes duplas; “Npc>2”, os nomes de partes múltiplas; “NpcM”, os nomes de partes massivas; e “NpcU”, os nomes de partes únicas.

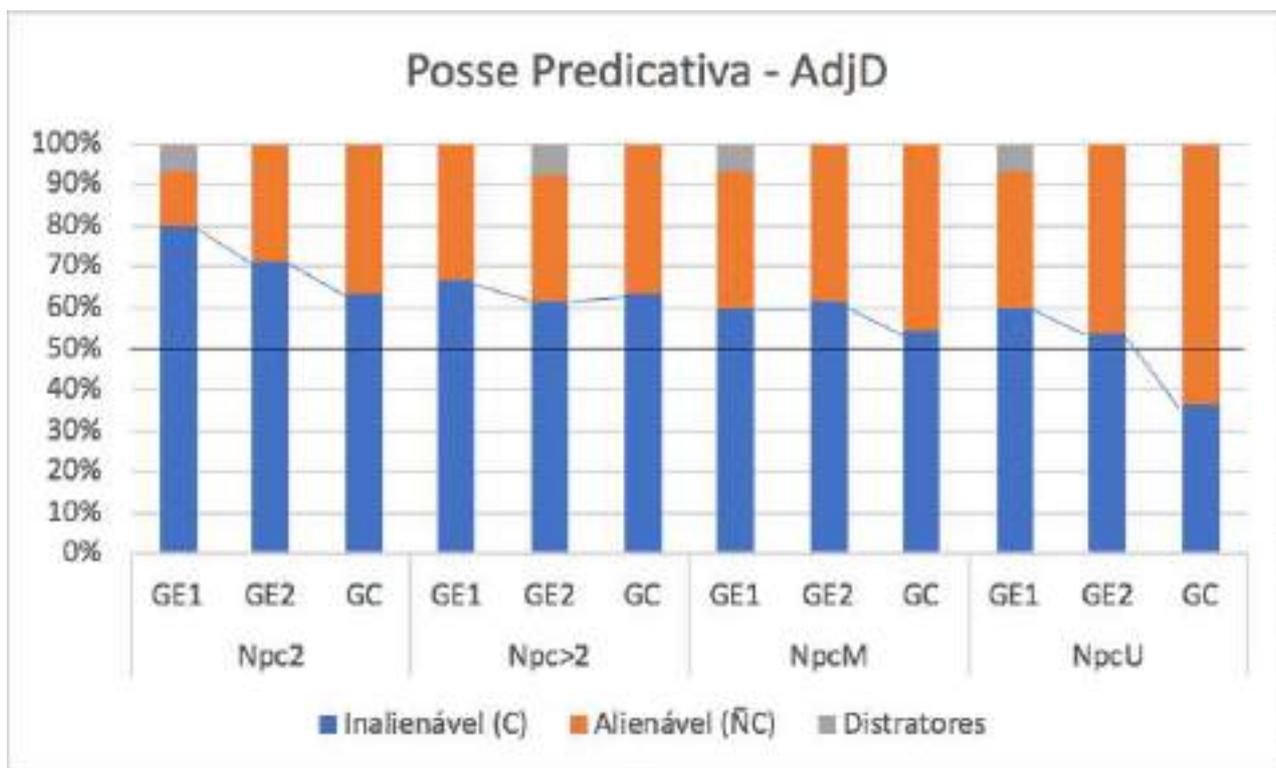
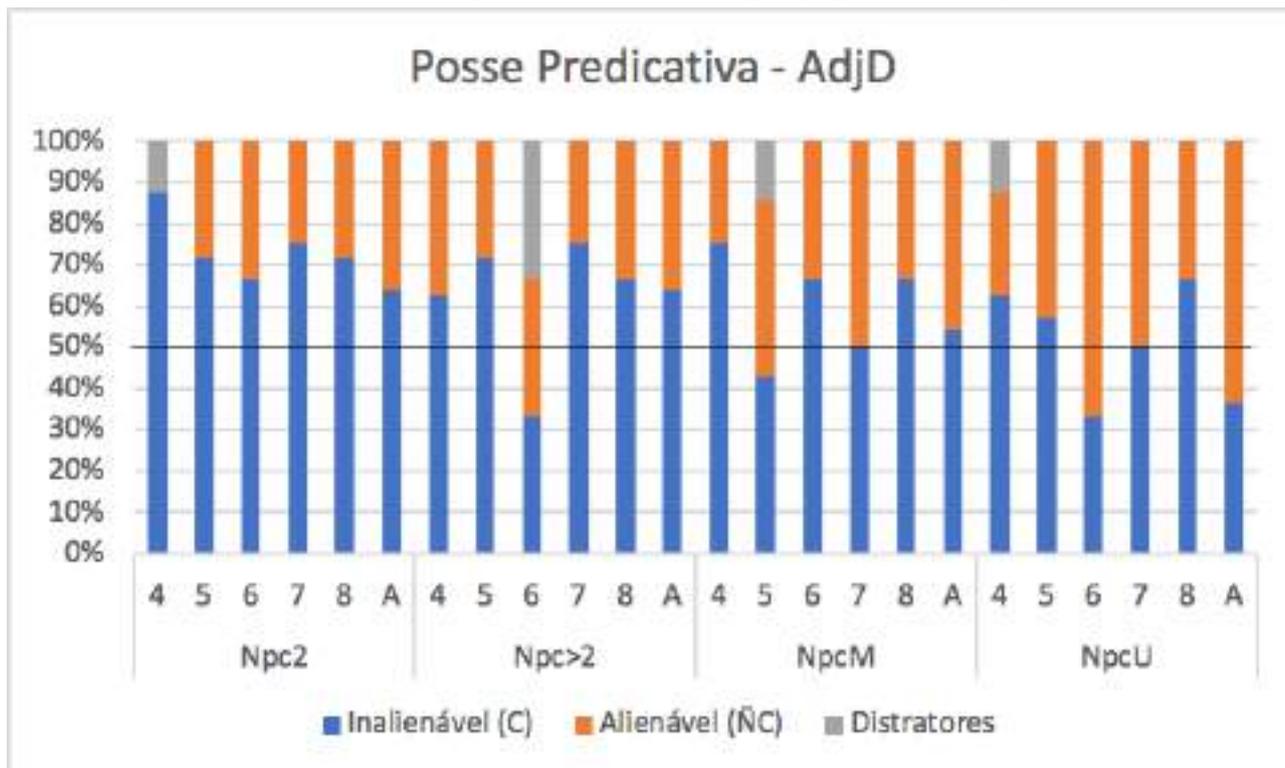
Observando o Gráfico 1 em detalhes, parece haver influência de duas variáveis, a saber, *Unicidade do nome inalienável* e *Tipo de adjetivo*. Por um lado, quando presentes na estrutura nomes inalienáveis não-únicos (duplos, múltiplos ou massivos), as taxas de preferência para a interpretação convergente – isto é, interpretação inalienável, no caso de presença de adjetivo na estrutura, e interpretação alienável, no caso em de ausência de adjetivo na estrutura – giram entre 45% e 63%. Sendo o valor mais alto referente às estruturas com adjetivos descritivos coocorrendo com nomes inalienáveis contáveis não-únicos. Por outro lado, quando estão presentes na estrutura nomes inalienáveis únicos, as taxas de preferência para a interpretação convergente são de 36% para estruturas com adjetivo e 18% para estruturas sem o adjetivo.

Assim, parece haver um sutil favorecimento da interpretação inalienável quando presentes na estrutura nomes inalienáveis contáveis e não-únicos e adjetivos descritivos, corroborando com a hipótese adotada. Além disso, parece ocorrer um favorecimento maior dessa mesma leitura quando, na estrutura, estão presentes nomes inalienáveis únicos e ausentes os adjetivos, contestando a hipótese aventada na pesquisa.

Já a leitura alienável parece ser favorecida quando estão presentes na estrutura nomes inalienáveis únicos coocorrendo com adjetivos, contrariamente à hipótese aventada. De uma forma mais sutil, pode-se dizer que essa leitura também parece ser favorecida quando presentes nomes inalienáveis múltiplos coocorrendo em estruturas sem adjetivo, tal qual prevê a hipótese adotada.

Em relação à gramática infantil, espera-se, de acordo com a hipótese tomada nessa pesquisa, uma preferência generalizada pela leitura inalienável, sendo esta a interpretação escolhida independentemente das condições sintáticas presentes na sentença. Apresentados em comparação com a gramática adulta, seguem abaixo os gráficos referentes aos resultados obtidos com relação à variável *Tipo de adjetivo* – adjetivo descritivo (AdjD), adjetivo restritivo (AdjR) e sem adjetivo (Sadj), respectivamente. Sendo que os Gráficos 2a, 3a e 4a, à esquerda, apresentam os resultados por idade – 4;0, 5;0, 6;0, 7;0, 8;0 anos de idade e A(dultos) –, enquanto os Gráficos 2b, 3b, e 4b, à direita, exibem os resultados por grupo etário – GE1, GE2 e GC.

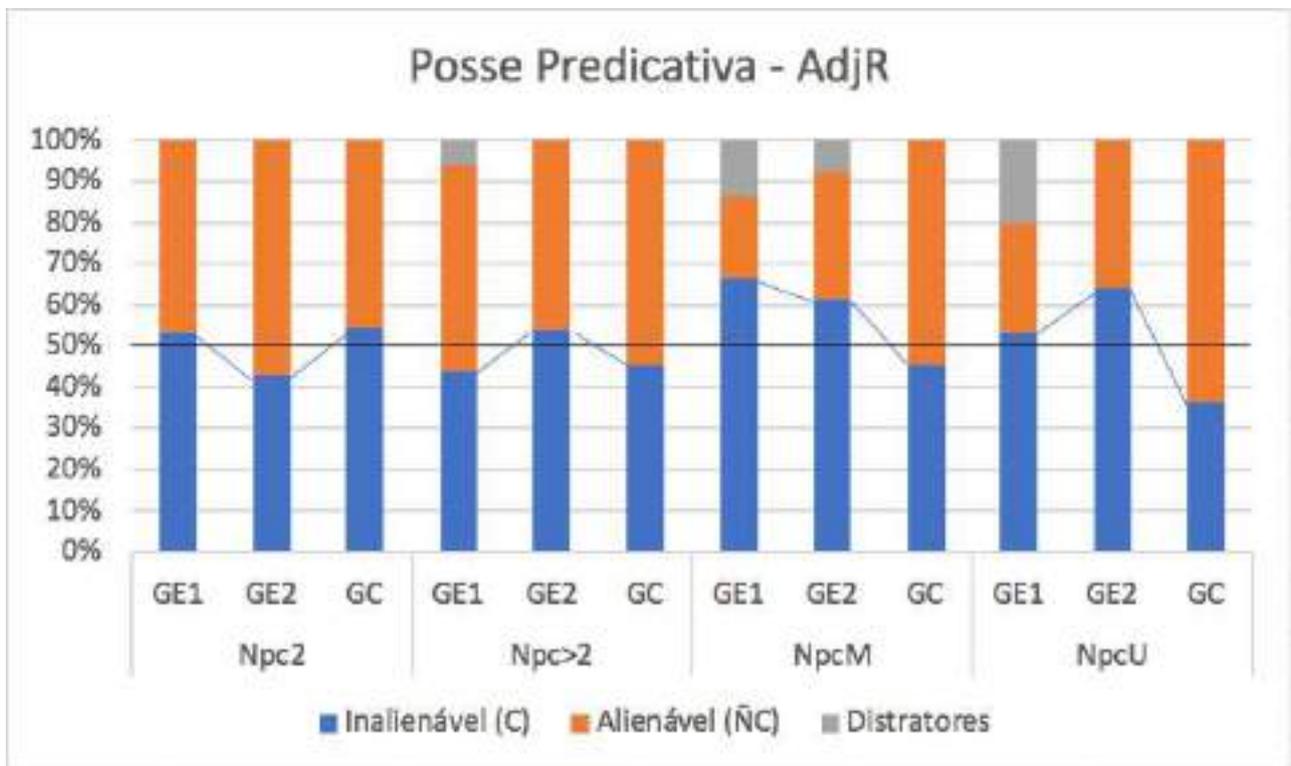
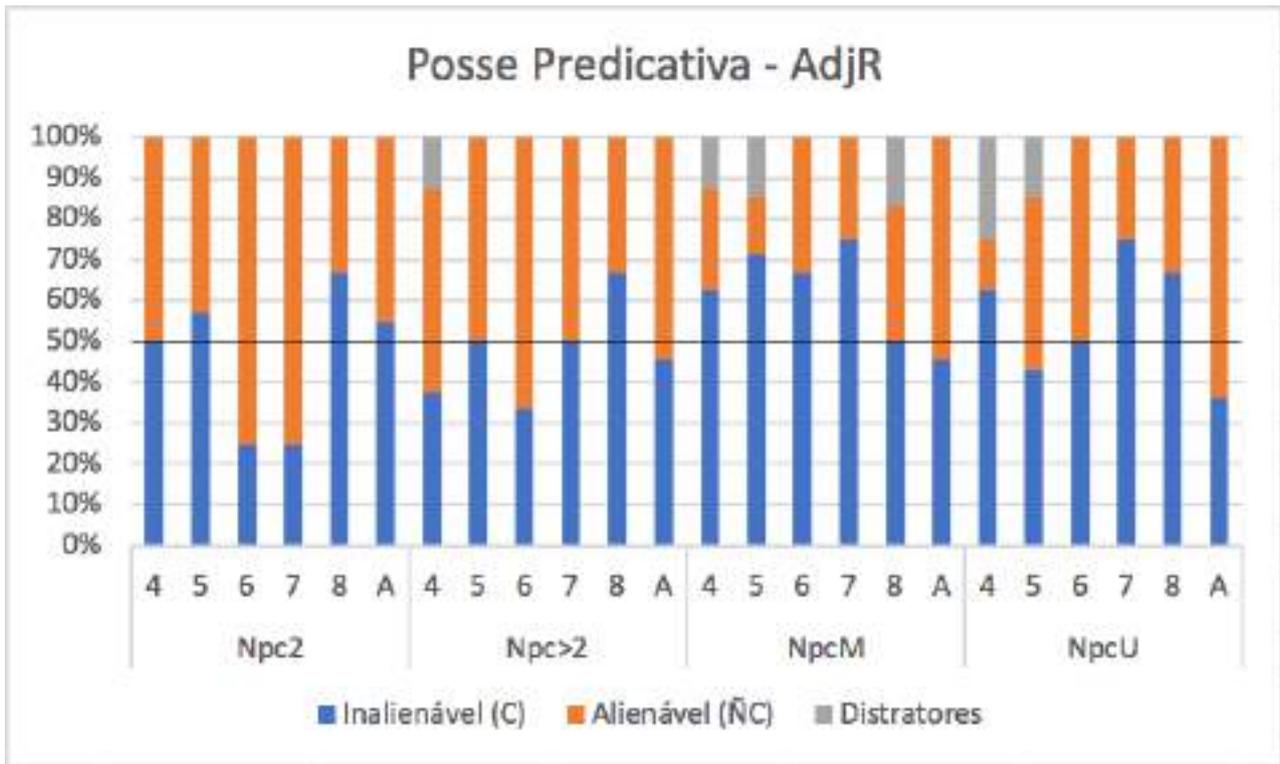
Gráfico 2a e 2b¹⁶ - Resultados por idade e Resultados por grupo etário – Adjetivo Descritivo



Fonte: elaboração própria

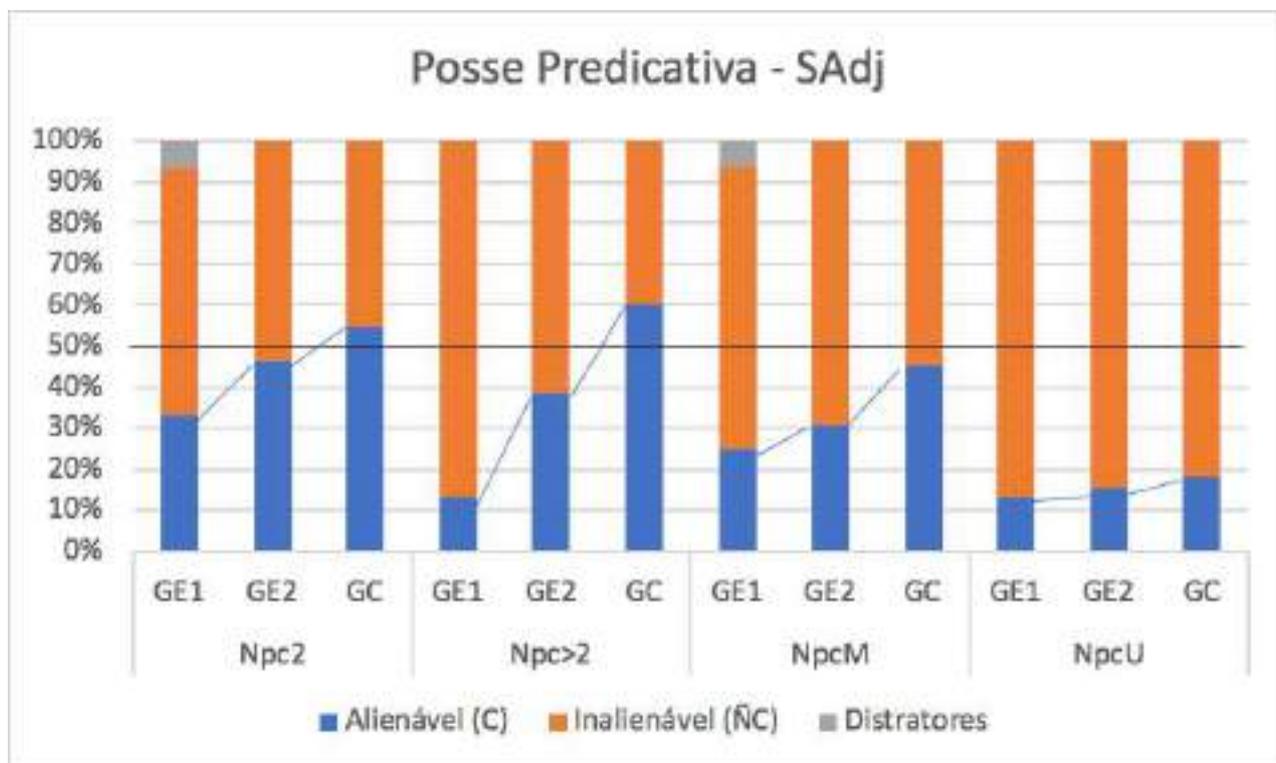
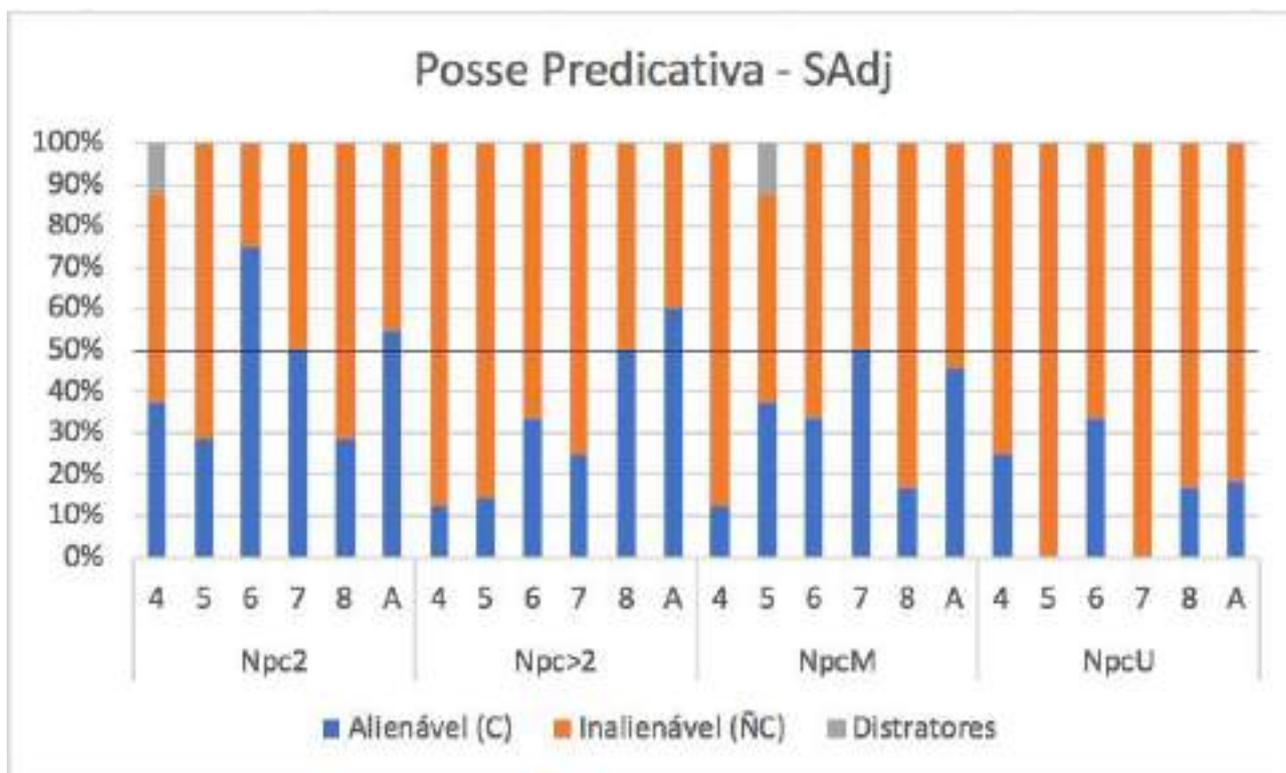
16 Nessa seção de gráficos, bem como na que será apresentada adiante, há, na legenda, além das leituras, também os distratores. Eles referem-se às taxas de escolha por imagens distratoras na tela do teste, sendo imagens que não veiculam nem a leitura alienável e nem a leitura inalienável.

Gráfico 3a e 3b - Resultados por idade e Resultados por grupo etário – Adjetivo Restritivo



Fonte: elaboração própria

Gráfico 4a e 4b - Resultados por idade e Resultados por grupo etário – Sem Adjetivo



Fonte: elaboração própria

O que pode ser observado nos Gráficos 2a-b e 3a-b e 4a-b acima é que, independentemente da presença e do tipo de adjetivo, isto é, seja o adjetivo descritivo ou restritivo ou ainda esteja esse elemento ausente, os participantes infantis, em especial aqueles pertencentes ao grupo dos mais jovens (GE1) tendem, de forma geral, a dar uma maior preferência à leitura inalienável, mesmo quando ela não seria a favorecida, haja vista a maior taxa em azul nos Gráficos 2b e 3b, assim como uma maior taxa de alaranjado no Gráfico 4b – seja pela hipótese adotada nesta pesquisa ou pela preferência exibida pelos resultados da gramática adulta. Os participantes infantis mais velhos (GE2), que também apresentam essa tendência, a exibem, porém, de uma forma mais moderada, quando comparados aos do GE1. O que indica uma aproximação em direção à gramática-alvo. Já os adultos, por outro lado, apresentam uma tendência ainda menor à leitura inalienável ou ainda uma tendência contrária – daí as curvas descendentes no Gráfico 2b, nos quais a leitura inalienável seria a esperada de acordo com a hipótese aventada e as curvas ascendentes no Gráfico 4b, em que a leitura alienável seria a esperada de acordo com a hipótese adotada.

Ainda que não haja uma curva ascendente ou descendente em todos os resultados da condição em que estão presentes adjetivos restritivos, apresentando-se em “U” ou “U invertido”, como pode-se observar no Gráfico 3b, as taxas exibidas pela gramática infantil se mostram, ao menos, semelhantes àquelas apresentadas pela gramática-alvo e/ou giram em torno de um comportamento *chance*.

Assim, pode-se dizer que a hipótese aventada nesta pesquisa, em relação ao processo de aquisição dessas construções possessivas parece ser corroborado. Estruturas possessivas inalienáveis adjetivas em construções de posse predicativa com o possuidor externo ao sintagma possuído apresentam, na gramática infantil do PB, uma preferência pela leitura inalienável mesmo sem qualquer adjetivo presente na estrutura do sintagma nominal possuído, tal como o fazem na presença de um adjetivo, aproximando-se da gramática adulta na medida em que avançam no processo de aquisição da linguagem.

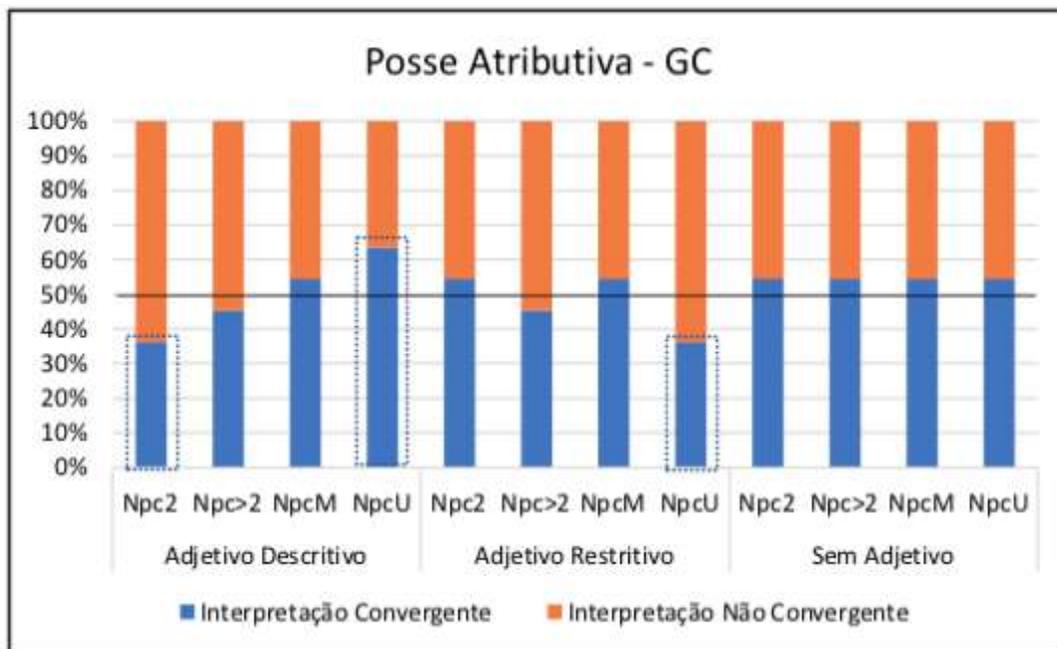
4.2. Posse atributiva

Em relação às hipóteses adotadas para a posse atributiva, seria previsto, por um lado, que a leitura inalienável fosse preferencial quando os adjetivos descritivos estivessem ausentes na estrutura, e, por outro lado, a preferência pela leitura alienável no caso da presença desses adjetivos na estrutura – caso o PB adulto se comportasse tal qual outras línguas românicas em relação a esse tipo de construção possessiva.

Entretanto, contrariamente ao esperado conforme a hipótese adotada, como ilustra o Gráfico 5 abaixo, a gramática adulta do PB parece apresentar, de forma geral, um comportamento *chance*

nessas construções, tal qual ocorre nas construções exibidas anteriormente. Observa-se uma sutil preferência um pouco mais marcada nos contextos em que está presente um adjetivo em coocorrência com um nome inalienável único ou duplo, como em (25a,d) e (26a,d).

Gráfico 5 - Resultados gerais do Grupo Controle em construções de Posse Atributiva



Fonte: elaboração própria

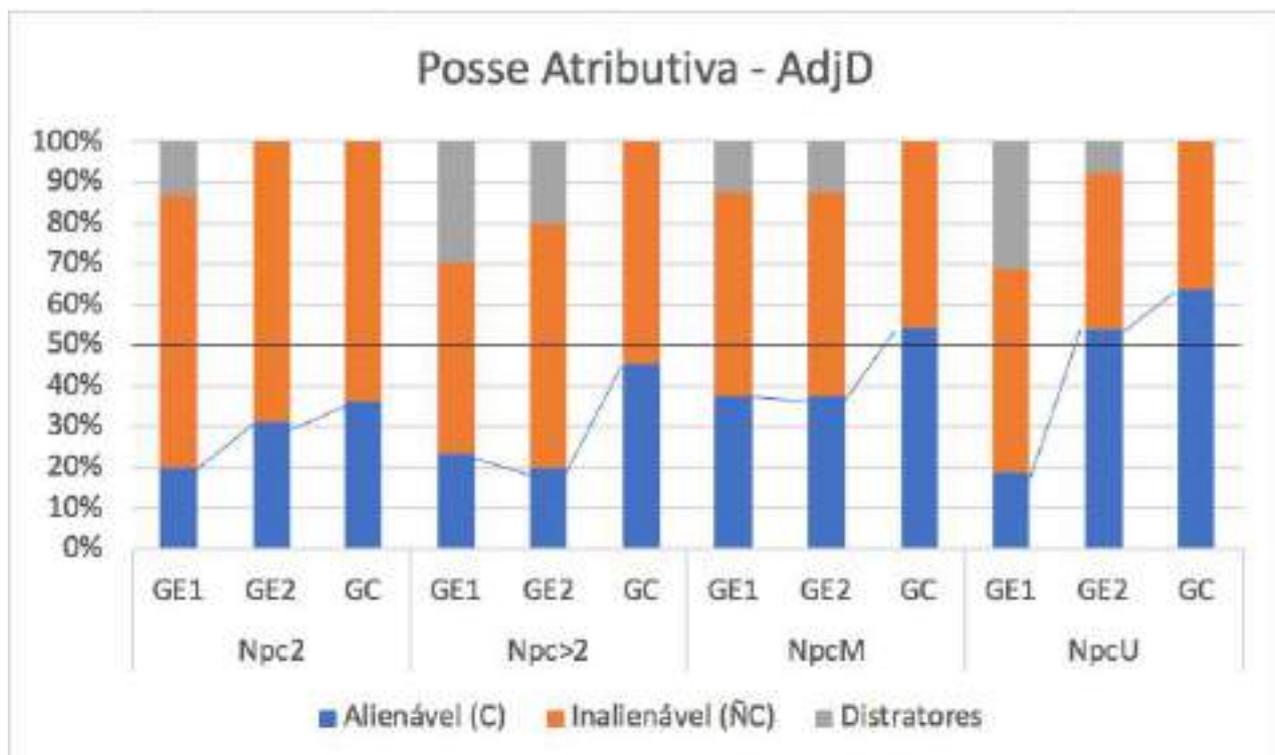
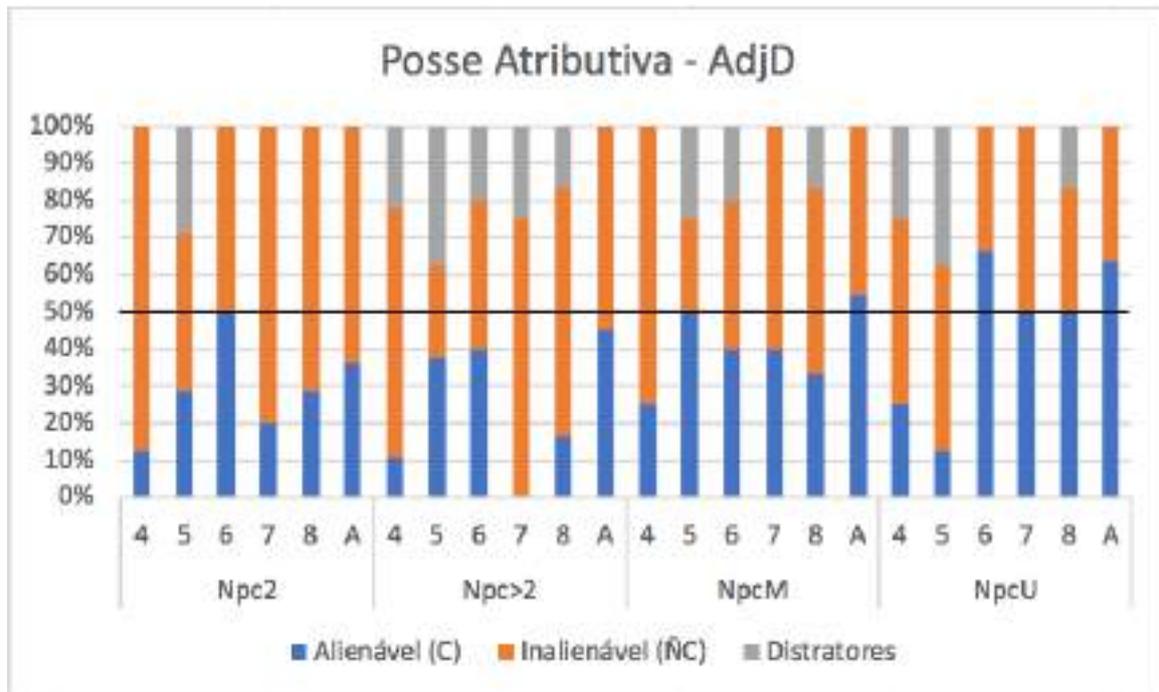
O Gráfico 5, de forma detalhada, também parece apresentar a influência das duas variáveis *Unicidade do nome inalienável* e *Tipo de adjetivo* – mais especificamente em relação à sua presença *versus* sua ausência. Em estruturas sem adjetivos observa-se um comportamento *chance*, independentemente da unicidade do nome inalienável. Já em estruturas com adjetivo, nota-se uma variação nas taxas de preferência da interpretação convergente – que, neste caso, seria a leitura alienável quando presente o adjetivo descritivo e, a leitura inalienável, quando presente o adjetivo restritivo – girando entre 63% para estruturas contendo nomes inalienáveis únicos coocorrendo com adjetivos descritivos e 36% para estruturas contendo nomes inalienáveis únicos coocorrendo com adjetivos restritivos e para estruturas contendo nomes inalienáveis duplos coocorrendo com adjetivos descritivos.

Assim, parece haver um favorecimento sutil, por um lado, da leitura alienável na coocorrência entre nomes inalienáveis únicos e adjetivos descritivos ou restritivos – sendo que apenas o primeiro caso corrobora com a hipótese aventada na pesquisa –, e, por outro lado, da leitura inalienável na coocorrência entre nomes inalienáveis duplos e adjetivos descritivos – um outro caso contrário à hipótese adotada.

Em relação à gramática infantil, em que se espera a superextensão da leitura inalienável para

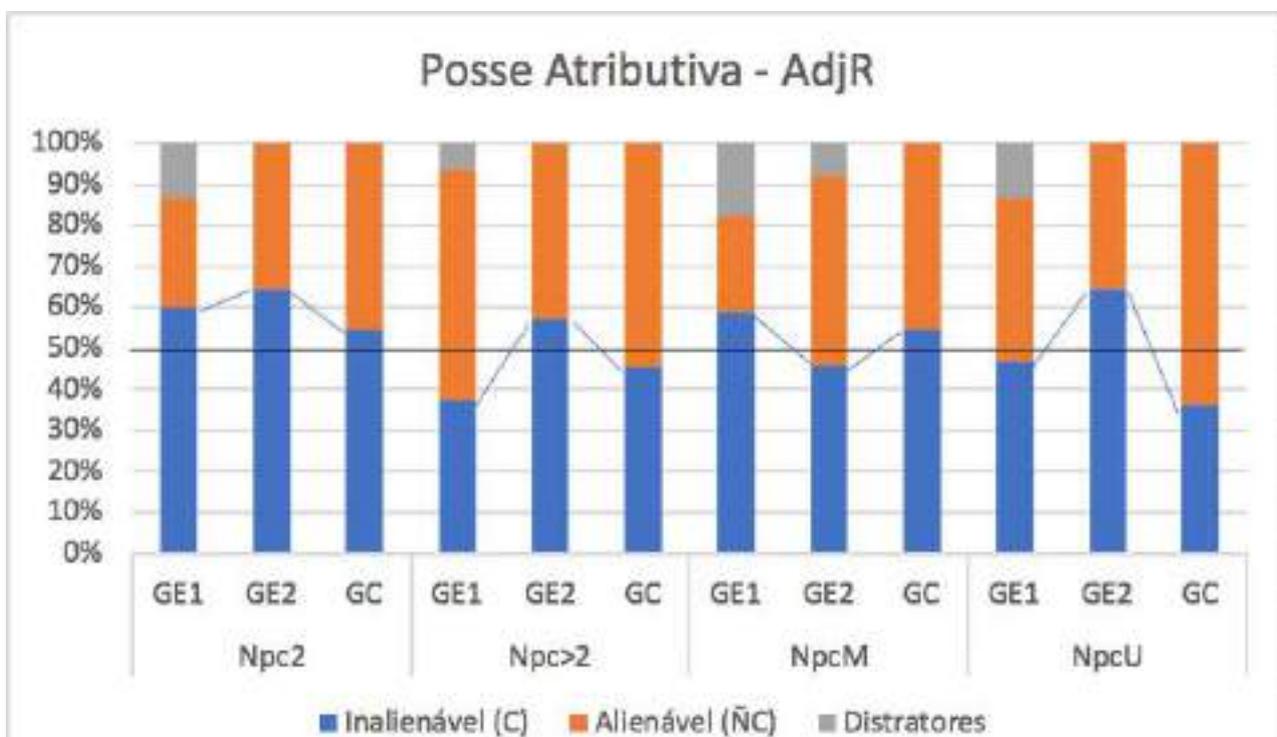
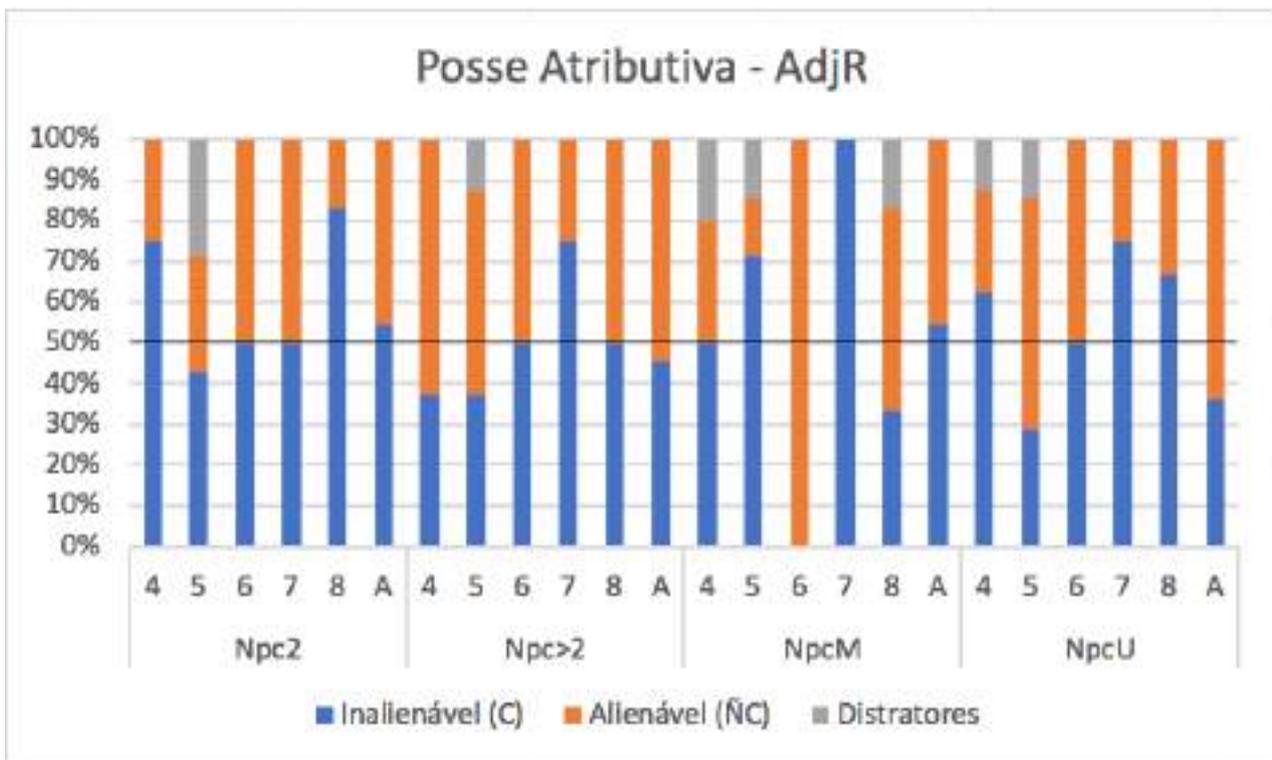
qualquer que seja a condição sintática apresentada, é, aqui também, exibida em comparação com a gramática adulta. Assim, seguem abaixo, organizados da mesma forma em que foram apresentados na subseção anterior, os gráficos referentes aos resultados obtidos.

Gráfico 6a e 6b - Resultados por idade e Resultados por grupo etário – Adjetivo Descritivo



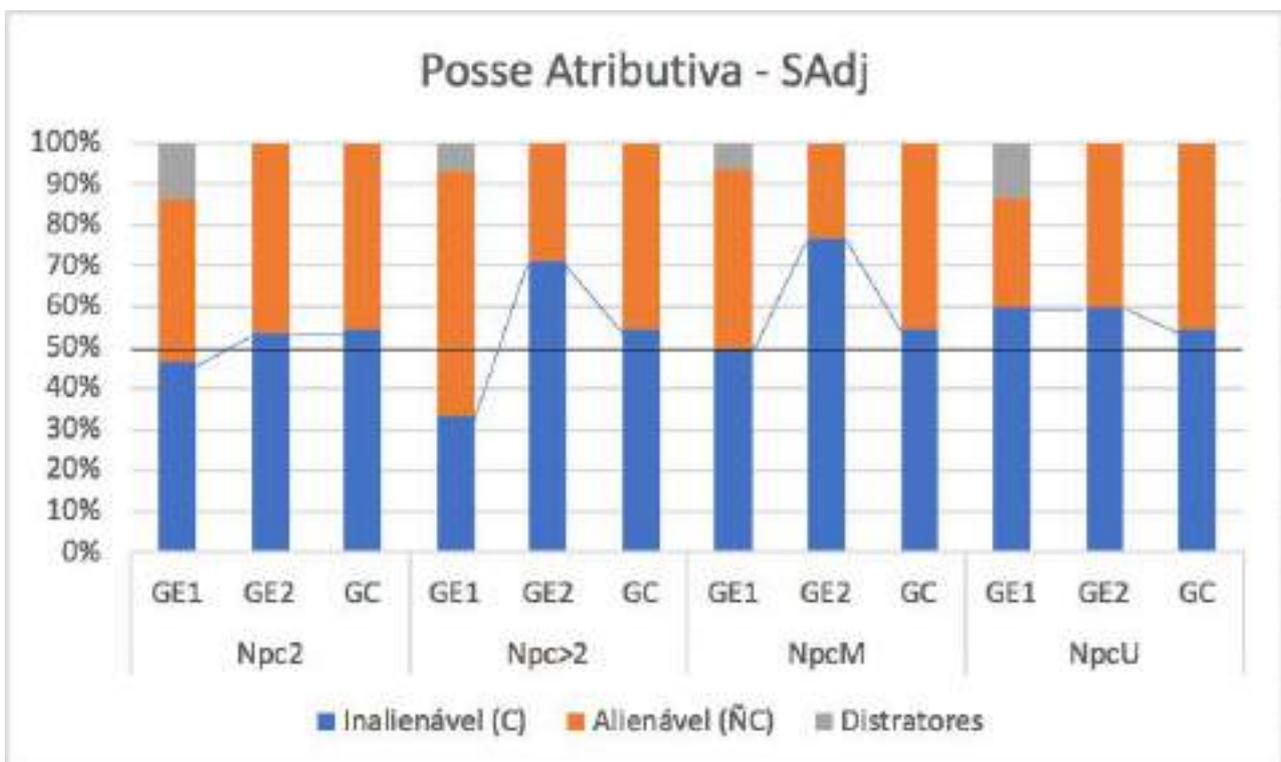
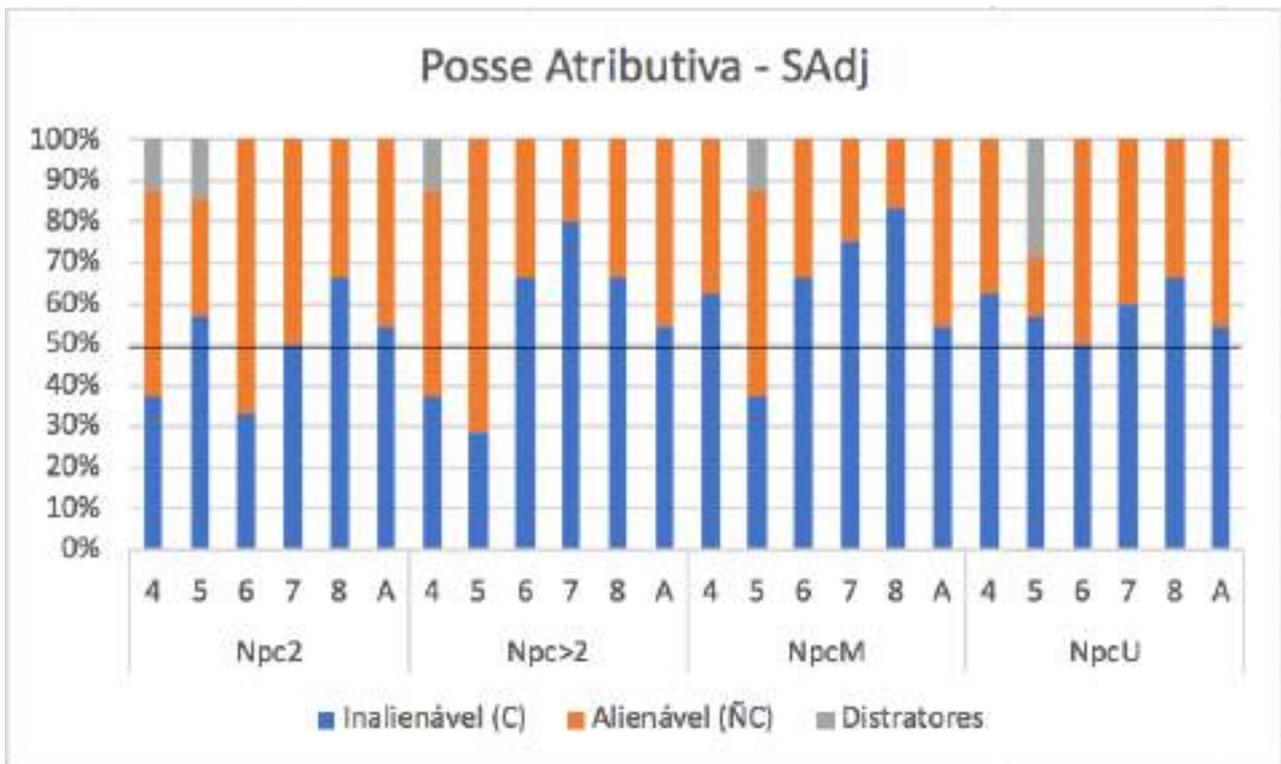
Fonte: elaboração própria

Gráfico 7a e 7b - Resultados por idade e Resultados por grupo etário – Adjetivo Restritivo



Fonte: elaboração própria

Gráfico 8a e 8b - Resultados por idade e Resultados por grupo etário – Sem Adjetivo



Fonte: elaboração própria

Os Gráficos 6a-b e 7a-b e 8a-b acima mostram que, de uma forma geral, parece haver uma

tendência mais sutil neste tipo de posse, comparando-se ao anterior, à escolha da leitura inalienável por parte dos participantes infantis. Destacam-se, em algumas condições, uma preferência mais marcada proveniente dos resultados do GE1 e, em outras, proveniente dos resultados do GE2. Em comparação aos resultados do GC, aqueles obtidos da gramática infantil parecem se destacar na preferência pela leitura inalienável, uma vez que as taxas em relação a essa leitura são, de forma geral, mais altas ou iguais àquelas obtidas dos resultados adultos. Em relação à gramática-alvo, observa-se no Gráfico 6b, que, por sua vez, trata de uma condição de restrição da leitura inalienável, curvas ascendentes em todos os contextos analisados. Esses resultados parecem mostrar um caminho de aquisição claro entre a gramática infantil e a gramática adulta. Já os Gráficos 7b e 8b, nos quais se esperava uma porcentagem maior de preferência em relação à leitura inalienável em ambos os grupos testados, parecem evidenciar uma variação entre as leituras disponíveis na gramática da língua além de padrões de curva variados.

Dessa forma, ao menos parcialmente, pode-se dizer que as hipóteses aventadas na pesquisa parecem ser corroboradas. Estruturas possessivas inalienáveis adjetivas em construções de posse atributiva com o possuidor externo ao sintagma possuído apresentam, na gramática infantil do PB, uma preferência pela leitura inalienável mesmo quando presente um adjetivo descritivo no sintagma nominal possuído, tal qual o fazem, ainda que em menor proporção, na presença de um adjetivo restritivo ou na ausência deste elemento na estrutura.

Considerações finais

Aprofundando os estudos realizados por Mendes (2015, 2017) em relação às estruturas possessivas inalienáveis, a presente pesquisa teve dois objetivos principais. Por um lado, buscou-se investigar o seu comportamento diante de restrições citadas na literatura da área referentes a construções predicativas e atributivas modificadas (ou não) por adjetivos (KAYNE, 1975; GUÉRON, 1985; VERGNAUD; ZUBIZARRETA, 1992), uma vez que, aparentemente, haveria variação no licenciamento da leitura inalienável também em PB. Por outro lado, examinou-se como se dá a aquisição dessas construções, especialmente no que concerne a essas restrições específicas.

Para tanto, foi desenvolvido e aplicado um experimento off-line a um total de 38 participantes, divididos em três grupos: GE1 (Grupo Experimental das crianças mais jovens); GE2 (Grupo Experimental das crianças mais velhas) e GC (Grupo Controle). O teste, baseado na Tarefa de Seleção de Figura (TSF), apresentou 24 sentenças-teste aos participantes – além de oito distratores – que resultaram da interação entre três variáveis: (i) construção possessiva (ii) unicidade do nome

inalienável e (iii) tipo de adjetivo.

Em relação aos resultados do GC, diferentemente do esperado de acordo com a hipótese adotada, a gramática adulta do PB não parece se comportar tal qual o francês em relação às restrições examinadas nesta pesquisa. Os dados revelam, de forma geral, um comportamento *chance*, apresentando, apenas sob poucas condições, algum tipo de preferência por alguma das duas leituras disponíveis. Acredita-se que isso se deva a um problema de ordem metodológica, uma vez que, diferentemente dos participantes infantis, os participantes adultos escolheram ambas as opções – alienável e inalienável – para responder às sentenças-teste.

Já em relação aos resultados dos GE1 e GE2, considerando-se que a leitura inalienável esteja disponível desde a fase inicial da gramática infantil e seja esta, inclusive, a interpretação generalizada esperada durante o início do processo de aquisição da linguagem, os resultados pertinentes a duas condições estruturais específicas apresentadas no experimento realizado requerem uma atenção especial, uma vez que são elas, de acordo com a hipótese adotada, que restringem a ocorrência desta leitura: (i) aqueles obtidos sob a condição de posse predicativa ocorrendo sem a presença de um adjetivo e (ii) aqueles obtidos sob a condição de posse atributiva ocorrendo com a presença de um adjetivo descritivo.

É possível observar, em ambos os casos, uma curva ascendente em relação à leitura preferida esperada (alienável), indicando, portanto, uma possível restrição da leitura inalienável – generalizada nas fases iniciais – ao longo do processo de aquisição da linguagem, aproximando a gramática infantil da gramática alvo na medida em que a criança cresce, conforme prevê a hipótese adotada.

Assim, entende-se que, embora a presente pesquisa colabore com o campo da Aquisição da Linguagem Experimental, sobretudo explorando as tão pouco estudadas construções possessivas inalienáveis, muito trabalho ainda precisa ser feito em relação a esse tema, especialmente dentro desse tipo de configuração estrutural, não apenas em relação à gramática infantil, mas também com respeito à gramática adulta.

REFERÊNCIAS

AUTHIER, Jean-Marc. *The syntax of unselective binding*. Los Angeles: University of Southern California, 1988.

AVELAR, Juanito. The comitative-copular basis of possessive-existencial constructions in Brazilian Portuguese. In: NUNES, Jairo. (ed.). *Minimalist essays on Brazilian Portuguese syntax*.

Linguistik Aktuell / Linguistics Today 142. Amsterdam / Philadelphia: John Benjamins, 2009, p. 139-160.

AVELAR, Juanito e CALLOU, Dinah. Sobre a emergência do verbo possessivo em contextos existenciais no português brasileiro. In: CASTILHO, Ataliba Teixeira de; TORRES MORAIS, Maria Aparecida Correa Ribeiro; LOPES, Ruth Elisabeth Vasconcellos; CYRINO, Sonia Maria Lazzarini. (orgs.). *Descrição, história e aquisição do português brasileiro*. São Paulo / Campinas: FAPESP / Pontes Editores, 2007, p. 375-402.

BARON, Irène e HERSLUND, Michael. Introduction. In: BARON, Irène; HERSLUND, Michael e SØRENSEN, Finn. *Dimensions of possession*. TSL 47. Amsterdam / Philadelphia: John Benjamins, 2001, p. 1-25.

CASTRO, Ana. *On possessives in Portuguese*. Lisboa / Paris: Universidade Nova de Lisboa / Université Paris 8 – Vincennes Saint-Denis, 2006.

CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.

_____. *Knowledge of language: its nature, origin and use*. Dordrecht: Foris, 1986.

FLORIPI, Simone e NUNES, Jairo. Movement and resumption in null possessor constructions in Brazilian Portuguese. In: NUNES, Jairo. (ed.). *Minimalist essays on Brazilian Portuguese syntax*. LA 142. Amsterdam / Philadelphia: John Benjamins, 2009, p. 51-68.

GERKEN, LouAnn e SHADY, Michelle. The picture selection task. In: MCDANIEL, Dana, MCKEE, Cecile e CAIRNS, Helen Smith. *Methods for assessing children's syntax*. Cambridge, MA / London: MITPress, 1996, p. 125-146.

GORDON, Peter. The truth-value judgment task. In: MCDANIEL, Dana, MCKEE, Cecile e CAIRNS, Helen Smith. *Methods for assessing children's syntax*. Cambridge, MA / London: MIT Press, 1996, p. 211-232.

GUÉRON, Jacqueline. Inalienable possession, PRO-inclusion and lexical chains. In: GUÉRON, Jacqueline; OBENAUER, Hans-Georg. e POLLOCK, Jean-Yves. (eds.). *Grammatical representation*. Dordrecht: Foris, 1985, p. 43-86.

HORNSTEIN, Norbert; ROSEN, Sara e URIAGEREKA, Juan. Integrals. UMWPL, 1995.

- KAYNE, Richard. *French syntax*. Cambridge, Massachusetts: The MIT Press, 1975.
- MENDES, Fernanda. Aquisição de estruturas possessivas inalienáveis: o caso dos nomes de parte do corpo em inglês americano e português brasileiro. In: *Revista de Estudos da Linguagem*, Belo Horizonte, v. 25, n. 3, 2017, p.1567-1611.
- MENDES, Fernanda. *Estruturas possessivas: a aquisição de posse inalienável no português brasileiro*. Campinas: UNICAMP, 2015.
- MENDES, Fernanda. *Estruturas possessivas: aquisição de posse funcional e posse inalienável no português brasileiro*. Florianópolis: UFSC, 2010a.
- MENDES, Fernanda. *A aquisição de estruturas inalienáveis adjetivas com verbos possessivos/ copulares em português brasileiro: um estudo experimental*. Manuscrito, 2010b.
- MENDES, Fernanda. *Estruturas inalienáveis adjetivas com verbos possessivos e copulativos em português brasileiro*. Manuscrito, 2010c.
- MITHUN, Marianne. The difference a category makes in the expression of possession and inalienability. In: BARON, Irène; HERSLUND, Michael e SØRENSEN, Finn. *Dimensions of possession*. TSL 47. Amsterdam / Philadelphia: John Benjamins, 2001, p. 285-310.
- MÜLLER, Ana Lúcia; NEGRÃO, Esmeralda Vailatti e NUNES-PEMBERTON, Gelza Matos. Adjetivos no português do Brasil: predicados, argumentos ou quantificadores? In.: ABAURRE, Maria Bernadete Marques e RODRIGUES, Angela (orgs.). *Gramática do português falado*. Vol. VIII. Campinas, SP: Editora da UNICAMP, 2002:317-344.
- MUNN, Alan, MILNER, Richard e SCHMITT, Cristina. Maximality and plurality in children's interpretations of definites. In: BAMMAN, David, MAGNITSKAIA, Tatiana e ZALLER, Colleen. (Eds.). *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville: Cascadilla Press, 2006:377-387.
- NEGRÃO, Esmeralda Vailatti; MÜLLER, Ana Lúcia; NUNES-PEMBERTON, Gelza Matos e FOLTRAN, Maria José. O adjetivo. In.: ILARI, Rodolfo e NEVES, Maria Helena de Moura. (orgs.). *Gramática do português culto falado no Brasil*. Vol. II. Campinas, SP: Editora da UNICAMP, 2008:371-402.
- PÉREZ-LEROUX, Ana Tereza; SCHMITT, Cristina e MUNN, Alan. Syntactic features and

discourse factors in children's interpretation of definite determiners in inalienable possessions. *Acts de l'ACL*, 2002a, p. 245-258.

PÉREZ-LEROUX, Ana Tereza; SCHMITT, Cristina e MUNN, Alan. The development of inalienable possession in English and Spanish. In: BOK-BENNEMA, Reineke; HOLLEBRANDSE, Bart; KAMPERS-MANHE, Brigitte e SLEEMAN, Petra. (Eds.). *Romance languages and linguistic theory*. Amsterdam: John Benjamins, 2002b, p. 199-216.

PÉREZ-LEROUX, Ana Tereza; SCHMITT, Cristina; MUNN, Alan. e DEIRISH, Michelle. *Learning definite determiners: genericity and definiteness in English and Spanish*. Boston University on Language Development, 2004.

PRIM, Cristina de Souza. *A sintaxe de adjetivos nas posições pré- e pós-nominal*. Florianópolis: UFSC, 2010.

RIBEIRO, Ilza. A formação dos tempos compostos: a evolução histórica das formas ter, haver e ser. In.: ROBERTS, Ian e KATO, Mary. (orgs.). *Português brasileiro: uma viagem diacrônica*. 2a edição. Campinas: Editora da UNICAMP, 1996, p. 343-386.

SCARPA, Ester Mirian. Aquisição da linguagem. In: MUSSALIM, Fernanda e BENTES, Anna Christina. *Introdução à lingüística: domínios e fronteiras*. Volume 2. São Paulo: Cortez, 2001, p. 203- 232.

SCHAEFFER, Jeannette. (Revisão). BAAUW, Sergio. Grammatical features and the acquisition of reference: a comparative study of Dutch and Spanish. *Glott International*, vol. 6, no 2-3, Feb-Mar 2002:65-71.

SCHAEFFER, Jeannette e MATHEWSON, Lisa. Grammar and pragmatics in the acquisition of article systems. *Natural Language and Linguistics Theory*, vol. 23, 2005, p. 53-101.

VERGNAUD, Jean-Roger e ZUBIZARRETA, Maria Luisa. The definite determiner and the inalienable constructions in French and in English. In: *Linguistic Inquiry*, v. 23, n. 4, Fall 1992, p. 595- 652

INTERFERÊNCIA DE VERBOS DE CONCORDÂNCIA DA LIBRAS (L1) NO USO DE PREPOSIÇÕES NA INTERLÍNGUA DE SURDOS APRENDIZES DE PORTUGUÊS L2 (ESCRITO)

INTERFERENCE OF AGREEING VERBS IN LIBRAS (L1) IN THE USE OF PREPOSITIONS IN THE INTERLANGUAGE OF DEAFS LEARNING L2 (WRITTEN) PORTUGUESE

Aline Mesquita¹

Heloisa Salles²

RESUMO

O estudo investiga a aquisição de português (escrito) como segunda língua (L2) por surdos falantes da Língua Brasileira de Sinais (Libras) como primeira língua, enfocando o uso da preposição ‘para’ introdutora do complemento dativo no Português Brasileiro (PB). Tomando por referência a hipótese do Acesso Parcial à Gramática Universal, segundo a qual existe interferência da L1 na aquisição de L2, e a hipótese de que o movimento direcional (DIR) em verbos de concordância na Libras e a preposição ‘para’ no PB são projeções de um núcleo relacional Q/P(\subseteq), que denota posse/inclusão, verificamos, por meio de estudo experimental, que a presença do morfema DIR, no verbo de concordância da Libras, favorece a ocorrência da preposição ‘para’ na interlíngua com o verbo correspondente no PB. A interferência é positiva nos casos de coincidência na marcação paramétrica entre as duas línguas, já que o verbo em português seleciona a preposição ‘para’ e seu correspondente em Libras manifesta DIR. No entanto, a interferência é negativa nos casos em que a preposição ‘para’ ocorre com verbos que não selecionam a preposição dativa no PB, e o verbo correspondente na Libras manifesta DIR.

Palavras-chave: Libras; dativo; aquisição de L2; preposições; verbos de concordância.

ABSTRACT

The study investigates the acquisition of (written) Portuguese as a second language (L2) by deaf people having the Brazilian Sign Language (Libras) as a first language (L1), focusing on the use of the preposition ‘para’ to introduce the dative complement in Brazilian Portuguese (BP). Taking into consideration the ‘Partial Access Hypothesis’, according to which there is interference of L1 in L2 acquisition, as well as the hypothesis that constructions with agreeing verbs in Libras and with the preposition ‘para’ in BP bear a relational head which denotes inclusion, namely Q/P(\subseteq), we noticed, by means of an experimental study, that the presence of DIR on the agreeing verb in Libras favors the occurrence of the preposition ‘para’ in the interlanguage. The interference is positive whenever there is coincidence in parametric marking between BP and Libras, as the verb selects the preposition ‘para’, and its correspondent in Libras manifests DIR. In turn the negative interference arises whenever the preposition ‘para’ occurs with verbs that do not take a dative preposition in BP while its correspondent in Libras manifests DIR.

Keywords: Libras; dative; L2 acquisition; prepositions; agreeing verbs.

1 Secretaria de Educação Profissional e Tecnológica. Ministério da Educação. E-mail: alinecrmesquita@gmail.com.

2 Instituto de Letras da Universidade de Brasília. E-mail: heloisasalles@gmail.com.

Introdução

O presente estudo investiga a aquisição da preposição dativa ‘para’ no Português Brasileiro (escrito), por surdos que têm a Libras³ como primeira língua (L1), considerando-se sua ocorrência na interlíngua com verbos bitransitivos e monotransitivos correspondentes aos verbos de concordância na Libras.⁴ Na análise, adotamos a Hipótese do Acesso Parcial à Gramática Universal (GU), em que se postula a interferência da L1 do aprendiz na aquisição da L2 (WHITE, 2003).

Para tanto, partimos da análise de verbos de concordância em Libras, em que um movimento direcional (DIR), na estrutura do sinal, relaciona o argumento externo e o argumento interno interpretado como alvo, propondo que é possível estabelecer uma relação de correspondência gramatical entre DIR e a preposição dativa (‘para’) em português. Essa correspondência, no entanto, não ocorre em todos os contextos: existem verbos de concordância em Libras que correspondem a verbos que não apresentam preposição dativa em PB; da mesma forma, sentenças com preposição dativa do PB podem corresponder a verbos simples da Libras (verbos que não apresentam movimento direcional). Essa relação será detalhada e exemplificada adiante.

A hipótese de trabalho é, portanto, a de que verbos de concordância em Libras interferem na aquisição das preposições dativas do PB, no sentido de favorecer o uso dessa categoria nos contextos em que há coincidência em relação ao uso de DIR e de ‘para’. A interferência será negativa nos casos em que o verbo de concordância corresponde a um verbo em português em que não ocorre a preposição dativa. Investigaremos, ainda, a relação dos verbos simples da Libras com os verbos correspondentes em PB que selecionam a preposição dativa. A previsão nesse caso é que haja interferência positiva, tendo em vista a relação entre a transitividade do verbo e a codificação formal do argumento interno marcado para o traço de ‘animacidade’ no sistema pronominal da Libras e pela preposição em português.

Considerando as questões apresentadas, analisaremos, inicialmente, as propriedades do complemento dativo nas línguas orais (LO) para, em seguida, verificarmos se essas propriedades podem ser observadas em verbos de concordância das línguas de sinais (LS). Finalizada essa etapa, analisaremos textos escritos em português por estudantes surdos, visando identificar a interferência

3 Adotamos a abreviatura ‘Libras’ para designar a Língua Brasileira de Sinais, conforme consta na legislação (Lei de Libras). Sabemos que a designação Língua de Sinais Brasileira e a respectiva sigla LSB é igualmente adotada.

4 O uso do rótulo ‘complemento dativo’ tem finalidade essencialmente descritiva, no sentido de distingui-lo de complementos sem preposição (objetos diretos) ou complementos não dativos introduzidos por outros tipos de preposição. Essa questão será retomada adiante, com a devida qualificação dos tipos de complementos envolvidos.

dos tipos de verbos em Libras (concordância e simples) na aquisição das preposições do PB. Os resultados permitem observar que, nas construções que apresentam verbos de concordância em Libras – bitransitivos ou monotransitivos –, há maior ocorrência da preposição ‘para’ nos dados da interlíngua. Essa preposição também está presente na interlíngua quando o verbo correspondente em Libras é simples e bitransitivo.

O artigo está dividido em 6 seções, incluída a introdução. Na seção seguinte (seção 2), detalharemos a relação entre o complemento dativo no PB e os verbos de concordância da Libras. Em seguida, na seção 3, examinaremos as características do complemento dativo nas línguas orais para verificarmos, na seção 4, se essas propriedades se aplicam às estruturas de concordância da Libras. Na seção 5, apresentaremos a metodologia adotada e os resultados da investigação da interlíngua dos surdos, que demonstram a interferência dos verbos de concordância na aquisição das preposições dativas do português. Por fim, na seção 6, apresentaremos as considerações finais.

2. A relação entre o complemento dativo no PB e os verbos de concordância em Libras

De acordo com a tradição gramatical, as construções dativas no PB padrão são aquelas iniciadas pela preposição ‘a’ e que podem ser substituídas pelo clítico ‘lhe’ nos contextos de 3ª pessoa (BECHARA, 2004) (cf. (1a-b) e (2a-b)). No PB vernacular, o complemento dativo é (preferencialmente) introduzido pela preposição ‘para’ (TORRES MORAIS; BERLINCK, 2007), sendo essa a preposição mais recorrente no *input* oferecido aos surdos, assim como a pronominalização pelo pronome tônico na estrutura do sintagma preposicional (cf. 1c e 2c):

- (1) a. Eu entreguei o livro ao menino.
b. Eu lhe entreguei o livro.
c. Eu entreguei o livro para o menino / para ele.
- (2) a. Eu telefonei ao meu amigo.
b. Eu lhe telefonei.
c. Eu telefonei para o menino/ para ele.

No âmbito da linguística gerativa, essas estruturas são amplamente estudadas, havendo diferentes propostas teóricas de análise do argumento dativo nas línguas. A discussão se detém na alternância dativa, pelo ponto de vista lexicalista ou construcionista (cf. KAYNE, 1984; MARANTZ, 1984, 1993; BAKER, 1988; LARSON, 1988; JACKENDOFF, 1990b; EMONDS, 1993; HALE; KEYSER, 1993; DEMONTE, 1995; PESETSKY, 1995; BORER, 1996; LEVIN; HOVAV, 1995; HARLEY, 2003; ANAGNOSTOPOULOU, 2003; PYLKKÄNEN, 2002; CUERVO, 2003; TORRES MORAIS; BERLINCK, 2007), para citar apenas alguns autores engajados nesse importante debate teórico e empírico.

O português brasileiro apresenta propriedades inovadoras em relação ao português europeu na realização do argumento dativo, destacando-se o estudo pioneiro de Ramos (1992), que identifica a perda da preposição ‘a’ nesse contexto sintático, relacionando-a à reanálise do sistema pronominal, com implicações para o surgimento da construção de objeto duplo (e por consequência da alternância dativa) no português brasileiro dialetal, ausente no português europeu e demais línguas românicas. As características inovadoras do argumento dativo no PB suscitam questões interessantes do ponto de vista da teoria de princípios e parâmetros, da mudança linguística e da análise formal das categorias envolvidas, particularmente em relação à natureza do caso oblíquo e ao papel das preposições e seus correlatos formais (cf. SALLES, 1997, TORRES-MORAIS; BERLINCK 2007; TORRES MORAIS; SALLES, 2010, entre muitos outros).⁵

Conforme observam Torres-Morais e Berlinck (2007), o complemento dativo apresenta propriedades que diferem entre as línguas.⁶ Esse argumento pode ocorrer com diversos tipos de verbos (transferência, criação, inacusativos, etc) e apresentar diferentes papéis temáticos (alvo, fonte, benefactivo, possuidor, etc), ocorrendo com verbos bitransitivos e monotransitivos, conforme ilustrado em (1) e (2), respectivamente. Interessantemente, as construções com complementos dativos são, em grande parte, realizadas, em Libras, por verbos de concordância (3 e 4)⁷:

(3) _{1s} ENTREGAR_{3s} LIVRO
‘Eu entreguei o livro para ele’.

(4) _{1s} TELEFONAR_{3s}
‘Eu telefonei para ele’.

5 Não é objetivo deste estudo aprofundar seja a questão teórica em relação às propriedades das estruturas com o argumento dativo, seja as condições que determinam o surgimento da gramática inovadora do PB, uma vez que a pergunta de pesquisa está voltada crucialmente para a análise da interferência da L1 na interlíngua de surdos aprendizes de português (escrito) como L2. Nesse sentido, optamos por um modelo de análise para qualificar formalmente as estruturas envolvidas, sabendo que existem modelos alternativos ou mesmo divergentes nessa discussão, particularmente em relação à perspectiva lexicalista ou construcionista.

6 Interessam ao presente estudo, indiretamente, as estruturas do português brasileiro padrão, por serem usadas na escrita. Além do acesso aos textos escritos (de caráter formal), o ensino de português como L2, no contexto educacional, propicia o acesso a estruturas vernaculares, no desenvolvimento das estruturas da língua. Não houve, neste estudo, o controle sistemático do *input* linguístico a que foram expostos os participantes do experimento. Parte-se da hipótese de que a escolarização é o meio pelo qual o surdo tem acesso às estruturas do português L2 (escrito). Estudos prévios antecipam a incidência de ‘para’ em detrimento de ‘a’ (cf. Mesquita 2008; Mesquita; Salles (2010); Salles *et al*, 2015) na escrita. A presente análise orientou-se no sentido de verificar a ocorrência da preposição ‘para’, que provou ser a categoria mais frequente, na comparação com outras preposições, nos contextos relevantes.

7 Para representar as sentenças da Libras, adotamos um sistema de transcrição que apresenta as seguintes propriedades: a) os sinais da Libras são representados por itens lexicais do português em letras maiúsculas, b) a apontação é representada por ‘IX’; c) as pessoas gramaticais do singular são representadas por 1s, 2s e 3s; d) a localização de um sinal no espaço está representada pelo sinal correspondente com uma letra ou número em subscrito que indica o *locus*. Ex: JOÃOa: João está associado ao ponto identificado como ‘a’; IX2: o sinalizador está apontando para a segunda pessoa (adaptado de FELIPE; MONTEIRO, 2007).

Os verbos de concordância em Libras, referidos originalmente por Ferreira-Brito (1995) como verbos direcionais ou flexionados, caracterizam-se por apresentarem um movimento direcional (M), marcando uma trajetória no espaço de sinalização, associada, no ponto inicial, ao sujeito e, no ponto final, ao objeto (alvo). Por hipótese, essa configuração codifica traços flexionais de concordância na estrutura do verbo (QUADROS, 1999; QUADROS; KARNOPP, 2004). Diferem, assim, dos verbos simples, que não apresentam movimento direcional associado aos argumentos do predicado, e dos verbos espaciais, que apresentam orientação para os argumentos locativos.

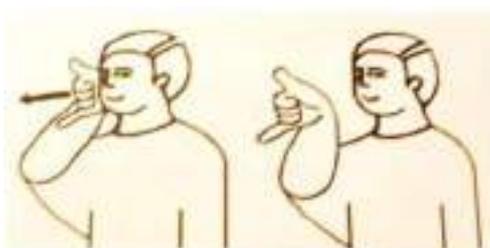
Observem-se as figuras 1 e 2, que representam os verbos ENTREGAR e TELEFONAR dos exemplos acima:

Figura 1: O sinal ENTREGAR em Libras



Fonte: Capovilla *et al.*, (2012, p.1086)

Figura 2: O verbo TELEFONAR em Libras



Fonte: Capovilla *et al.* (2012, p. 2342)

A correspondência entre as línguas⁸, apresentada nos exemplos acima, nem sempre ocorre: há casos em que a sentença com complemento dativo no PB é realizada por um verbo simples da Libras, e há verbos de concordância que não correspondem a uma sentença com complemento dativo no PB. Os quadros abaixo ilustram as relações possíveis entre tipo de verbo em Libras (concordância x simples) e tipo de complemento em PB (preposição dativa x preposição não dativa x sem preposição)

⁸ Neste ponto, cabe frisar que esta análise considera apenas complementos dativos argumentais – excluindo, portanto, sintagmas *benefactivos*. Desconsideram-se também os sintagmas locativos, associados a verbos espaciais em Libras. A complexidade dessas relações está no fato de que as línguas mapeiam as relações argumentais de formas distintas. É o caso, por exemplo, do verbo ‘ajudar’, que seleciona complemento marcado como acusativo em português (pela ausência da preposição), e como dativo em alemão (pela marca do caso dativo na estrutura do DP). Essa diferença translinguística pode ser explicada em termos da forma como a língua realiza a representação do predicado. Conforme sugerem Manzini e Franco (2016), a partir de uma sugestão de Svenonius (2002, citado pelos autores), o contraste paramétrico pode ser expresso na estrutura de complementação, pois uma dada língua pode ser sensível (ou não) à estrutura de subeventos/ estados do predicado. Dessa forma, segundo os autores, a opção positiva do parâmetro seria aquela em que essa relação é marcada por uma categoria relacional, o morfema de caso dativo no alemão. Estendemos essa análise ao morfema DIR em Libras. Essa questão será retomada adiante.

nas duas línguas. No quadro 1, está ilustrada a relação de verbos bitransitivos em Libras e em PB: o grupo I apresenta verbos que selecionam complemento dativo em PB e correspondem a verbos de concordância em Libras, enquanto no grupo II, a relação é entre verbos que selecionam dativo em PB e verbos simples em Libras. Já no quadro 2, está ilustrada a relação de verbos monotransitivos em PB e em Libras: o grupo I apresenta os verbos já especificados anteriormente; o grupo III apresenta verbos que não selecionam dativo em PB e verbos de concordância em Libras; e o grupo IV apresenta verbos que não selecionam dativo em PB e verbos simples em Libras. Cabe notar que os verbos exemplificados nos quadros são encontrados em ambas as línguas⁹.

Quadro 1: A relação dativo-concordância com verbos bitransitivos em PB e Libras.

GRUPO	Português	LIBRAS	Verbos
I	Dativo	Concordância	avisar, contar, dar, entregar, ensinar, enviar, informar, mandar, mostrar, perguntar, responder, vender, pagar.
II	Dativo	Simples	escrever, explicar, falar, dizer, prometer

Fonte: Elaborado pelas autoras

Quadro 2: A relação dativo-concordância com verbos monotransitivos em PB e Libras.

GRUPO	Português	LIBRAS	Verbos
I	Dativo	Concordância	Telefuntar, ouvir, obedecer
III	Não-dativo	Concordância	A) Preposicionados (no PB): zombar de, cuidar de, bater em, atirar em, mandar em, jogar em B) Não-preposicionados (no PB): abandonar, abençoar, acusar, ajudar, demita, desprezar, demonstrar, influenciar, proteger, provocar, vencer, ver A) Preposicionados (no PB): persuadir em, gostar de, precisar de, concordar com, confiar em B) Não-preposicionados (no PB): amar, comer, beber, ter, saber.
IV	Não-dativo	Simples	conhecer

Fonte: Elaborado pelas autoras

⁹ Alguns verbos em português são realizados por um único sinal em Libras ('dar' e 'entregar'; 'falar' e 'dizer') – essa questão não é relevante para nosso estudo, pois o foco é discutir o fenômeno em função da transitividade dos verbos e do tipo de complemento selecionado. Agradecemos ao parecerista anônimo por apontar a necessidade de esclarecer a existência de formas-significados e usos entre os verbos em PB e em Libras. Trata-se de uma questão complexa que se resolve no nível da análise léxico-conceitual e que vai além do escopo da presente discussão.

Apresentamos, a seguir, exemplos com verbos de cada grupo ilustrado nos quadros, sentenças (3) e (4) (repetidas abaixo), representando o grupo I (dativo-concordância); sentença (5), representando o grupo II (dativo – simples); sentenças (6) e (7), representando o grupo III (não dativo – concordância); sentenças (8) e (9), representando o grupo IV (não dativo – simples):

- | | |
|------------------------------------|----------------------------------|
| (3) $_{1S}$ ENTREGAR $_{3S}$ LIVRO | ‘Eu entreguei o livro para ele’. |
| (4) $_{1S}$ TELEFONAR $_{3S}$ | ‘Eu telefonei para ele’. |
| (5) IX_1 ESCREVER CARTA IX_3 . | ‘Eu escrevi uma carta para ele’. |
| (6) $_{1S}$ ZOMBAR $_{2S}$ | ‘Eu zombei de você’. |
| (7) $_{1S}$ AJUDAR $_{2S}$ | ‘Eu ajudei você’. |
| (8) IX_1 PENSAR IX_2 | ‘Eu penso em você’. |
| (9) IX_1 AMAR IX_2 | ‘Eu amo você’. |

Conforme mencionado, a hipótese de trabalho é a de que verbos de concordância interferem positivamente na aquisição das preposições presentes nas estruturas dativas do PB. Esses verbos interferem negativamente, quando o verbo correspondente em português seleciona um complemento não dativo. Quanto aos verbos simples, quando seu correspondente em PB seleciona dativo, consideramos que há interferência positiva, tendo em vista seu estatuto bitransitivo, definido pela relação entre sua estrutura léxico-conceptual e a codificação do traço de animacidade do argumento alvo no sistema pronominal (a ser detalhada adiante).

Assumindo a hipótese da Gramática Universal, conforme formulada no quadro teórico gerativista (cf. CHOMSKY, 1995), analisaremos as propriedades formais do complemento dativo de verbos bitransitivos e monotransitivos nas LOs, para verificar a correspondência estrutural em relação aos verbos de concordância das LS (bitransitivos e monotransitivos).

3. O complemento dativo nas línguas orais

Segundo os estudos gerativos, particularmente Chomsky (1986; 1995), o dativo é um Caso inerente, assim designado pela relação temática entre o sintagma determinante (DP) e o núcleo sintático que o seleciona – seja verbal ou nominal. A abordagem que adotamos neste trabalho é a de Manzini e Franco (2016), que assume a correspondência direta entre léxico e sintaxe – um tipo de lexicalismo, com a possibilidade de projeção das propriedades do léxico como sistemas de traços formais articulados por uma teoria de eventos, como no caso da abordagem de Hale e Keyser (1993), admitindo-se igualmente modelos de análise, como o de Ramchand (2008), em que os argumentos são introduzidos por núcleos sintáticos que mapeiam a estrutura do evento, e por categorias lexicais

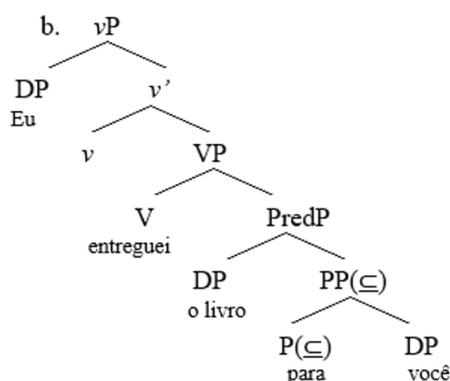
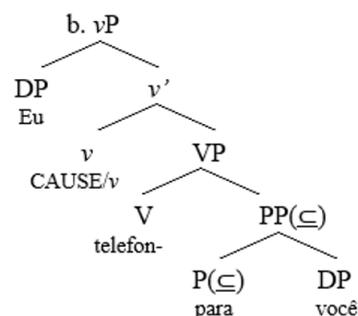
(como preposições), presentes na estrutura interna do predicado.

De acordo com Manzini e Franco (2016), o Caso oblíquo engloba os Casos dativo e genitivo¹⁰, que se caracteriza por denotar uma relação entre dois argumentos, representada por uma noção de inclusividade, identificada pela notação \subseteq . O complemento dativo se identifica pela presença de uma categoria marcadora do argumento alvo, que pode ser uma preposição, nas línguas românicas, ou algum item do léxico, como a flexão de caso no nome, em línguas como o alemão e o latim. Nessa proposta, portanto, tanto a marcação morfológica no sistema pronominal (Q), quanto a preposição (P) realizam a relação de inclusividade como um núcleo sintático (\subseteq), que introduz a relação entre dois argumentos internos (possuído e possuidor). Assim, os complementos dativos são introduzidos na sentença pelo núcleo $P(\subseteq)/Q(\subseteq)$.

O núcleo $P/Q(\subseteq)$ é responsável por representar tanto os dativos presentes em estruturas bitransitivas quanto em estruturas monotransitivas. A relação estabelecida por $Q(\subseteq)$ nos verbos bitransitivos se dá entre o argumento de $Q(\subseteq)$ (alvo) e o argumento tema presente na estrutura do sintagma verbal (VP). No caso dos verbos monotransitivos, os autores propõem que a relação seja entre o argumento dativo (alvo) e um subevento do predicado, este último entendido como uma estrutura que compreende um nível verbal (verbo leve) e um nível nominal. Segundo essa hipótese, os predicados transitivos são bi-eventivos e podem ser analisados como uma estrutura projetada a partir de um nome marcado para um traço eventivo. Essa intuição foi formalizada na análise de Hale e Keyser (1993) e também por Chomsky (1995), por meio da categoria funcional ν , na estrutura do predicado transitivo, que tem o papel de introdutor do argumento externo e de licenciador do argumento interno. Assim, a sentença “Ele me telefonou” teria a estrutura mais abstrata do tipo “Ele me deu um telefonema”.

Adotamos a análise de Manzini e Franco (2016) para os complementos dativos no PB. Dessa forma, o núcleo $P/Q(\subseteq)$ é lexicalizado pela preposição ‘para’ (ou ‘a’, mais restrita à escrita) em verbos bitransitivos e monotransitivos. Apresentamos, a seguir, as estruturas das construções dativas bitransitivas (10) e monotransitivas (11):

¹⁰ A abordagem que adotamos neste trabalho distingue-se da proposta do núcleo aplicativo, entendido como a categoria funcional introdutora do complemento dativo nas línguas (MARANTZ, 1993; PYLKANNEN, 2002). A hipótese do núcleo aplicativo faz uma distinção entre complementos dativos (introduzidos pelo núcleo aplicativo) e oblíquos (introduzidos pela preposição), que conforme explicado acima, não levamos em consideração.

(10) a. Eu entreguei o livro para você.(11) a. Eu telefonei para você.

Na seção a seguir, examinamos análises prévias dos verbos de concordância nas línguas de sinais, além da possibilidade de analisá-los como projeções do núcleo relacional.

4. Os verbos de concordância nas línguas de sinais

Tendo em vista a análise referente ao complemento dativo nas línguas orais, passamos à investigação acerca da estrutura sintática dos verbos de concordância nas línguas de sinais (particularmente a Libras), uma vez que as construções que apresentam esses verbos em Libras podem corresponder a construções com complementos dativos do PB.

Verbos de concordância nas LS – assim como verbos que selecionam dativo nas LOs – podem selecionar um ou dois argumentos internos. Considerando-se a semelhança no mapeamento sintático do argumento interno alvo em ambos os predicados, cabe indagar: é possível atribuir a mesma estrutura sintática na configuração bitransitiva e monotransitiva dos verbos das línguas de sinais? Observem-se os exemplos (12) e (13):

(12) a. _{1S} ENTREGAR _{2S} LIVRO
b. ‘Eu entreguei o livro para você’.

(13) a. _{1S} AJUDAR _{2S}
b. ‘Eu ajudei você’.

Uma questão que se coloca é se existe diferença na estrutura sintática dos verbos em Libras (12) e (13), a depender da transitividade. Alguns estudos atribuem a mesma análise aos argumentos internos de verbos bitransitivos e monotransitivos (PADDEN, 1983; MEIR, 2002 e RATHMANN; MATHUR, 2002); outros afirmam que são complementos com estruturas sintáticas diferentes (JANIS, 1995; QUADROS; QUER, 2008, 2010). Iniciemos com as abordagens que dão tratamento idêntico a

esses complementos¹¹.

Padden (1983) propõe que verbos de concordância bitransitivos e monotransitivos estabelecem a concordância com o sujeito e com o objeto direto final de forma idêntica. A autora adota o quadro teórico da gramática relacional (PERLMUTTER, 1979 *apud* PADDEN, 1983), para explicar a concordância na Língua de Sinais Americana (ASL). Segundo essa abordagem, uma sentença consiste em uma série de elementos linguísticos, na relação gramatical que eles geram (predicado, sujeito, objeto direto, objeto indireto e relações oblíquas) e no nível em que ocorrem essas relações (inicial e final).

Para explicar a concordância verbal em ASL, Padden (1983) propõe dois modelos de análise. Na primeira, *uni-level analysis*, é estabelecida a regra: o verbo (P) concorda com sujeito (1) e objeto direto (2), caso não haja objeto indireto (3); se houver, o verbo concorda com 1 e 3. A concordância com 1 e 2 seria referente aos verbos monotransitivos (cf. 15), enquanto a concordância com 1 e 3 seria referente aos verbos bitransitivos (cf. 14):

(14) P 1S(ENTREGAR)2S – 1(1S) – 2 (LIVRO) – 3 (2S)

(15) P 1S(AJUDAR)2S – 1 (1S) – 2 (2S)

No segundo modelo de análise, *advancement analysis*, a autora postula uma estrutura com dois níveis (estrato inicial e estrato final) para representar as sentenças bitransitivas. No estrato inicial, é mantida a relação P (predicado) – 1 (sujeito) – 2 (OD) – 3 (OI). Já no estrato final, a relação seria P – (predicado) – 1 (sujeito) – *chômeur* – 2 (OD).¹² A regra para a concordância seria, portanto, simplificada: o verbo concorda com 1 e 2 final. Dessa forma, a estrutura de verbos bitransitivos e monotransitivos seria unificada, pois, em ambos, a concordância se daria sempre com 2 final (cf. 16 e 17):

(16) Estrato inicial: P 1S(ENTREGAR)2S – 1(1S) – 2 (LIVRO) – 3 (2S)

Estrato final: P 1S(ENTREGAR)2S – 1(1S) – *chômeur* (LIVRO) – 2 (2S)

(17) P 1S(AJUDAR)2S – 1 (1S) – 2 (2S)

Meir (2002) também defende a hipótese de que os argumentos internos dos verbos de concordância

11 Não são muito numerosos os estudos que discutem contrastivamente a estrutura sintática de verbos de concordância bitransitivos e monostransitivos nas línguas de sinais. Por esse motivo, apresentamos, neste artigo, a discussão sobre as propriedades desses verbos, sem considerar (ou debater) diferenças nos pressupostos teóricos de cada abordagem, limitando-nos a demonstrar que as análises convergem na identificação do movimento como uma categoria gramatical (e não apenas um parâmetro da estrutura do sinal).

12 Na gramática relacional, o argumento ‘*chômeur*’ é assim designado por corresponder a uma posição sintática não preenchida, que se mantém implícita (ou que foi destituída de sua posição original e movida para uma posição periférica).

bitransitivos e monotransitivos devem receber a mesma análise, adotando uma abordagem baseada na estrutura argumental. Examinado a Língua de Sinais Israelense (ISL), a autora argumenta que os verbos de concordância são formados pela configuração de mão, que corresponde à raiz verbal, por um morfema direcional (DIR), responsável pela marcação de papel temático, e pela orientação da mão, que marca a função sintática. O morfema DIR corresponde ao movimento da trajetória do verbo, que parte do argumento fonte em direção ao argumento alvo. Dessa forma, a autora unifica a presença desse morfema com morfemas das línguas orais como as preposições, observando que, como as marcas de caso nas LOs, DIR, nas LS, pode ser realizado independentemente ou como um morfema preso.

O que diferencia as duas modalidades é que, enquanto nas LOs, se observa a linearidade entre o morfema e a raiz, nas LS, DIR ocorre simultaneamente ao verbo. Ao se afixar a um verbo em LS, DIR define o movimento de sua trajetória em relação aos argumentos dispostos em uma relação semântica fixa: fonte-alvo. O parâmetro da orientação da mão (ponto para o qual a palma da mão – ou as pontas dos dedos – está voltada, é analisado como o marcador de Caso dativo, uma vez que esse parâmetro está sempre voltado para o *locus* em que se situa o argumento interno. As LS, portanto, introduzem o argumento interno com marca de Caso dativo na estrutura do sinal, enquanto o sujeito nominativo não teria nenhuma marca morfológica.

A mesma análise é proposta para os verbos reversos (como ‘pegar’), assim denominados por apresentarem a chamada ‘concordância reversa’: o ponto inicial do movimento é realizado no ponto que corresponde ao objeto, e o ponto final, no *locus* que representa o sujeito. De acordo com Meir (2002), na realização de verbos reversos, o movimento se mantém fonte-alvo, uma vez que o movimento parte do objeto e pelo sujeito respectivamente), e a orientação da mão se mantém voltada para o argumento interno. Assim, de acordo com a autora, os complementos dos verbos de concordância bitransitivos e monotransitivos são formados pelos mesmos elementos e, conseqüentemente, devem apresentar as mesmas funções: semanticamente, a relação fonte-alvo (marcado por DIR) e, sintaticamente, o caso dativo (marcado pela orientação da mão)¹³. Como esses verbos apresentam as mesmas propriedades (DIR e orientação da mão), seus argumentos não devem ser analisados de forma diferente.

Na mesma vertente, Rathmann e Mathur (2002) argumentam que os argumentos internos que participam da concordância nas LS compartilham a mesma posição estrutural em verbos bitransitivos e monotransitivos, o que se aplica também aos argumentos de verbos simples (cf. 18, ilustrado por

¹³ Empregamos o termo ‘caso’ com letra minúscula, conforme adotado por Meir (2002), uma vez que se refere à marcação morfológica.

dados da Libras em 19):

- (18) a. NP (sujeito) V NP (OD inanimado)
- b. NP (sujeito) V NP (OD animado)
- c. NP (sujeito) V NP (OI animado) NP (OD inanimado)

- (19) a. IX₁ ESTUDAR MATEMÁTICA
- b. _{1S}AJUDAR_{2S}
- c. _{1S}DAR_{2S} LIVRO

De acordo com essa análise, que leva em consideração a animacidade dos argumentos, os objetos diretos inanimados de (18a) e (18c) ('matemática' e 'livro') compartilham a mesma posição, tanto na estrutura do verbo simples ('estudar') quanto na estrutura do verbo de concordância ('ajudar' e 'dar'). Da mesma forma, o OD animado na estrutura do verbo em (18b)/(19b) deve ser realizado na mesma posição estrutural do OI na estrutura do verbo em (18c)/(19c) (os pronomes de segunda pessoa, glosados como 2S). Segundo Rathmann e Mathur (2002), se essa análise estiver correta, a concordância nas LS seria caracterizada em termos de posição estrutural.

Passemos, agora, às abordagens que dão tratamentos sintáticos distintos aos complementos de verbos de concordância. Iniciemos com a proposta de Janis (1995), que parte da distinção entre objeto direto e objeto indireto. Nessa análise, a autora propõe uma hierarquia que considera casos, funções sintáticas e papéis temáticos para definir o argumento controlador da concordância. Deste modo, o objeto indireto tem preferência como controlador de concordância, seguido do objeto direto e, por último, do sujeito. Janis (1995), portanto, distingue as funções de objeto direto e objeto indireto – este, presente na estrutura do verbo bitransitivo (assim como o objeto direto); aquele, presente na estrutura do verbo monotransitivo.

Quadros e Quer (2008, 2010) também atribuem propriedades sintáticas aos complementos dos verbos de concordância monotransitivos e bitransitivos: objeto direto e objeto indireto, respectivamente. Essa análise apoia-se na relação entre papel temático e função sintática do argumento interno: o argumento alvo corresponde ao objeto indireto, e o argumento tema, ao objeto direto. Segundo Quadros e Quer (2008, 2010), uma das fragilidades da análise de Meir (2002) é a generalização de que o movimento da trajetória (DIR) é sempre fonte-alvo. Para os autores, essa correlação não se sustenta, uma vez que, nos verbos monotransitivos, o argumento interno tem o papel temático de tema – sendo portanto, um argumento acusativo.

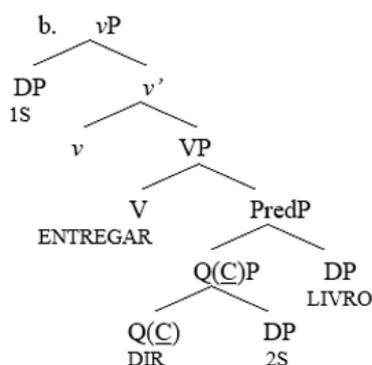
Outro exemplo, segundo Quadros e Quer (2008), de que o movimento da trajetória não é fonte-alvo está nas construções que apresentam verbos auxiliares em libras. Tais verbos têm a função

de realizar a concordância, no nível morfológico, quando unido a um verbo, pelo movimento, cuja trajetória parte do sujeito em direção ao objeto. Quadros e Quer (2008) observam que, nesse caso, se trata do mesmo morfema direcional (DIR) postulado por Meir (2002), com a diferença de que, quando ocorre com a função de verbo auxiliar, não é realizado simultaneamente ao verbo, mas linearmente. A questão é que o auxiliar apresenta o movimento da trajetória do sujeito ao objeto, mesmo se o verbo a que se liga é do tipo de reverso, em que a orientação é fonte-alvo (cf. 20 com o verbo reverso PEGAR). Assim, segundo Quadros e Quer, não é possível manter a análise unificada de Meir, uma vez que a padronização do movimento (fonte-alvo) estipulada pela autora não se mantém em todos os casos.

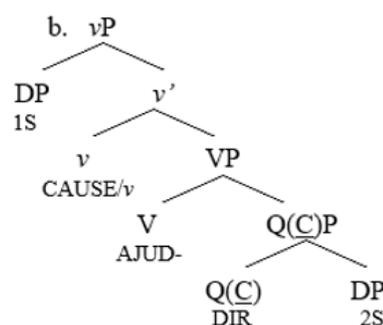
(20) IXx IXy xAUXy (y)PEGAR
 ‘X pegou Y’

Os estudos citados apresentam divergências em relação ao enquadramento teórico ou às propriedades postuladas. No entanto, pode-se dizer que as análises convergem na identificação do movimento direcional como uma categoria gramatical, e não apenas um parâmetro da estrutura do sinal. Apesar de reconhecer que os papéis temáticos são distintos, adotamos a hipótese de que verbos de concordância bitransitivos e monotransitivos apresentam a mesma estrutura sintática. Em particular, seguimos a análise de Manzini e Franco (2016) em relação a complementos dativos nas LOs, e assumimos que verbos de concordância bitransitivos e monotransitivos em LSs são projeções da categoria relacional P/Q (\subseteq), lexicalizada pelo movimento direcional (morfema DIR) (cf. 21 e 22), em oposição a verbos simples monotransitivos, que não apresentam tal núcleo (cf. 23):

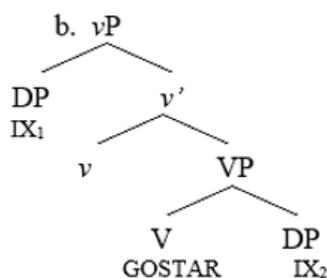
(21) a. 1sENTREGAR_{2s} LIVRO (‘Eu entreguei o livro para você’).



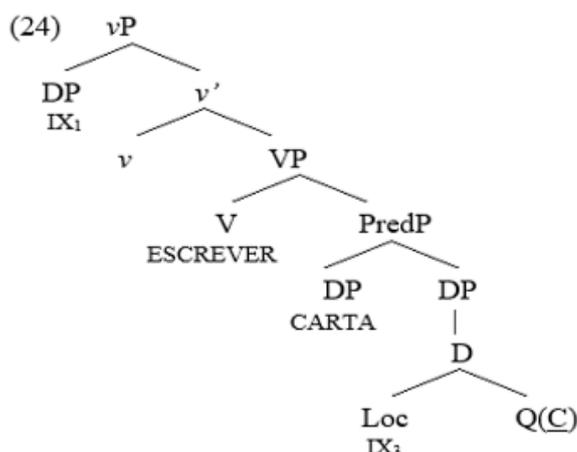
(22) a. 1sAJUDAR_{2s} (‘Eu ajudei você’)



(23) a. IX₁ GOSTAR IX₂ (‘Eu gosto de você’)



Resta, ainda, explicar a estrutura das sentenças com verbos simples bitransitivos como IX₁ ESCREVER CARTA IX₃ ('Eu escrevi uma carta para ele'). Por se tratar de um predicado bitransitivo, propomos que a relação entre os dois argumentos internos é mediada pela categoria relacional, exatamente como no caso do verbo de concordância, embora, nesse caso, a presença do núcleo relacional esteja associada ao argumento alvo determinado na estrutura léxico-conceitual do verbo, o que implica a ocorrência de uma construção mediada por Q(C).



A hipótese de que o núcleo de inclusividade está presente na estrutura do sintagma determinante (DP) tem como correlato a codificação do caso dativo no sistema pronominal, em diferentes línguas (como 'lhe', no português, e 'le', no espanhol) – também associado ao traço de animacidade. Embora o sistema pronominal da Libras não inclua um contraste desse tipo, consideramos a possibilidade de que o sinal de apontação (IX), nessa estrutura, tenha uma realização marcada.¹⁴

5. Examinando a interlíngua dos surdos

Nesta seção, apresentamos o resultado parcial (com adaptações) de teste¹⁵ com os estudantes surdos, realizado por Mesquita (2019), para verificar a interferência dos verbos de concordância da Libras na aquisição das estruturas dativas do português.

14 O traço [+animado] é inerente ao pronome de 1ª e 2ª pessoa, em oposição à 3ª pessoa (BENVENISTE, 1966). Na análise de Manzini e Franco (2016), a presença do operador de transferência de posse (material ou cognitiva), na estrutura do predicado, atribui o traço relacional (⊆) ao argumento alvo, gramaticalizando o traço [+animado]. A possibilidade de o sistema pronominal da Libras apresentar marcação distintiva com esses verbos (seja na realização da apontação, por um movimento mais alongado, seja por marcação não manual (direção do olhar), ou ainda pelo uso do movimento orientado para argumento alvo na realização do verbo) nos foi sugerida por Ronice Quadros (c.p.). Essa hipótese requer estudo mais aprofundado.

15 Os testes foram submetidos e aprovados pelo Comitê de Ética em Pesquisa, do Instituto de Ciências Humanas (CEP/IH) da Universidade de Brasília (Parecer número 3.036.218).

5.1 Metodologia

5.1.1 Participantes

Participaram do experimento 29 surdos, com idades entre 21 e 37 anos, congênitos (ou ensurdecidos antes dos três anos de idade), bilaterais, com grau de surdez profundo ou severo. Todos os estudantes são usuários de Libras. O primeiro contato com essa língua se deu fora do ambiente familiar, uma vez que seus pais são ouvintes¹⁶.

O teste foi realizado em duas escolas distintas – que denominaremos escola A e escola B. Conforme amplamente referido nos estudos surdos, a escola é o contexto primordial de acesso à língua oral (na modalidade escrita) pelo surdo. Disso decorre uma situação de bilinguismo peculiar, em que a LS é a primeira língua – pelo imperativo da adequação à condição perceptual dessas pessoas – e a LO é desenvolvida primordialmente no contexto educacional, na modalidade escrita, por ser essa modalidade igualmente a mais adequada.¹⁷

Os participantes do experimento apresentam os seguintes níveis de proficiência do PB: iniciante (8 estudantes), básico (10 estudantes) e intermediário (11 estudantes). Os informantes estão identificados pelas iniciais de seu nome, acrescido do número que corresponde ao nível de proficiência: 1 (iniciante), 2 (básico), 3 (intermediário).

O mesmo teste foi aplicado a um grupo de controle formado por 10 ouvintes, com idade entre 18 e 35 anos. Os resultados em relação ao grupo controle não são discutidos neste estudo, cabendo ressaltar apenas que a previsão do teste mostrou-se compatível com a produção nativa.

5.1.2 Materiais

O experimento consistiu em um teste semi-estruturado de produção eliciada, em que os participantes deveriam escrever uma frase de acordo com a imagem, utilizando as palavras escritas

16 Para a maioria dos participantes, o contato com a Libras se deu até a adolescência (somente três participantes adquiriram a Libras na fase adulta). Optamos por manter todos os participantes na pesquisa e não controlamos essa variável na análise, embora saibamos que o período de aquisição de L1 tem significado importante, tendo em vista a hipótese do período crítico (Lennenberg, 1967). Consideramos como variável principal o grau de surdez (profundo ou severo). A Libras é tida como a L1 dos participantes da pesquisa, uma vez que a modalidade visual é a mais adequada ao surdo e todos os participantes se consideram proficientes na língua. Agradecemos ao parecerista que nos ressaltou a importância do período de aquisição da Libras. Consideraremos essa questão em estudo futuro.

17 Dessa situação decorre um caso especial de bilinguismo, no qual o uso do português (escrito) como segunda língua é funcionalmente restrito (mas inevitável, no contexto da escolarização), enquanto a L1 assim se define pelo imperativo da condição perceptual do falante, mesmo que não corresponda à situação (ideal) de acesso na primeira infância. São inúmeros os estudos surdos que qualificam essa situação, destacando-se Skliar (1996). Em relação ao bilinguismo dos surdos, considere-se a contribuição de Quadros (1997), Fernandes (2003), entre muitos outros.

em cada balão (cf. figuras 3 e 4).¹⁸ Essas palavras referiam-se ao argumento externo, aos argumentos internos e aos verbos contidos nos grupos ilustrados no quadro 3 abaixo:

Quadro 3: Verbos utilizados no experimento.

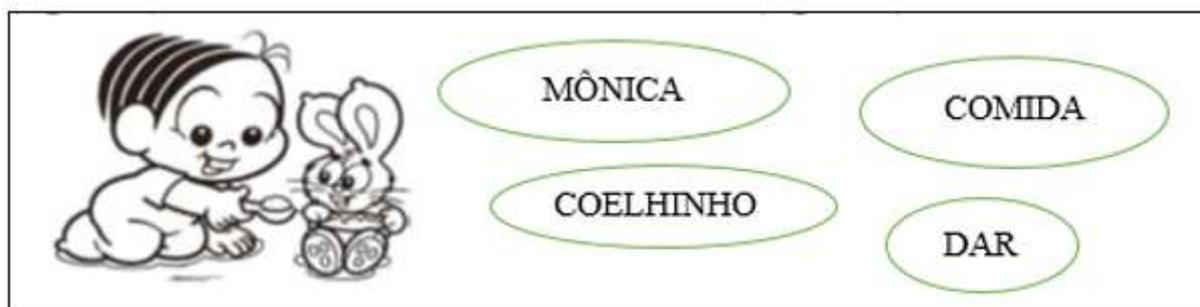
GRUPO	PB	LIBRAS	Verbos testados
I	Dativo	Concordância	dar, ensinar, responder, contar, telefonar
II	Dativo	Simples	escrever, explicar, falar, dizer, prometer
IIIA	Preposicionado	Concordância	zombar de, cuidar de, bater em, atirar em, mandar em
IIIB	Não-preposicionado	Concordância	ajudar, acusar, abandonar, ver, vencer
IVA	Preposicionado	Simples	pensar em, gostar de, precisar de, concordar com, confiar em
IVB	Não-preposicionado	Simples	amar, comer beber, ter, saber

Fonte: elaboração própria.

O Quadro 3 ilustra a relação entre tipo de complemento no PB (dativo e não dativo) e tipo de verbo em Libras (concordância e simples), que já foi descrita nos quadros 1 e 2 (seção 2). No quadro 3, estão exemplificados apenas os verbos usados no experimento: 5 verbos de cada grupo, que resultou em um total de 30 sentenças por participante.

Abaixo, apresentamos dois exemplos das imagens utilizadas, uma com verbo bitransitivo (figura 3) e outra com verbo monotransitivo (figura 4):

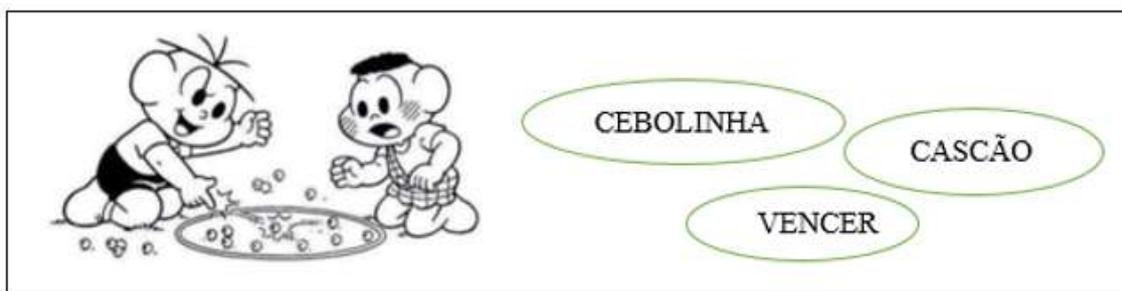
Figura 3: Imagem usada no experimento – verbo DAR (cf. Colorir.org).



Fonte: elaboração própria

¹⁸ Na pesquisa de Mesquita (2019), foram realizados mais dois testes (julgamento de gramaticalidade e preenchimento de lacunas), que não constam neste artigo devido à limitação de espaço.

Figura 4: Imagem usada no experimento – verbo VENCER (cf. Colorir.org).



Fonte: elaboração própria

5.1.3 Procedimentos

O teste foi realizado em duas escolas distintas – escola A e escola B, como afirmado anteriormente. A escola A, mantida pela Secretaria de Estado de Educação do Distrito Federal – SEEDF – fornece atendimento a alunos surdos, bem como a professores, intérpretes e comunidade interessada.

Antes da aplicação do teste, foi entregue aos participantes uma lista com os verbos do experimento e o respectivo sinal em Libras. A pesquisadora leu cada verbo com os participantes para se certificar de que todos conheciam a palavra em português e o sinal em Libras. Alguns alunos desconheciam o verbo ‘zombar’, em português, mas conseguiram identificar o seu sentido pelo sinal correspondente em Libras. No que se refere aos sinais em Libras, todos os participantes conheciam os sinais ilustrados.¹⁹

Também foi solicitado aos participantes que informassem à pesquisadora se algum daqueles verbos poderia ter um sinal diferente do ilustrado. A nossa preocupação era identificar se algum verbo, analisado como sendo de concordância em Libras, também poderia ter como sinal variante um verbo simples (e vice-versa) – o que ocorreu com os verbos simples ‘explicar’, ‘falar’ e ‘dizer’. Alguns participantes apresentaram para esses verbos sinais que apresentam movimento direcional – o que muda o seu *status* para verbo de concordância.

Concluída essa etapa, iniciamos o teste. O enunciado foi apresentado em Libras, pela pesquisadora, aos participantes.

¹⁹ Em resposta ao questionamento de um parecerista anônimo, informamos que não houve discussão com os participantes em relação às propriedades gramaticais dos verbos da Libras e do português, abordados no estudo, no sentido de estabelecer identificação entre o parâmetro do movimento e a preposição ‘para’. O objetivo era investigar um estado da interlíngua em relação ao uso (ou não) da preposição nos contextos relevantes – e não o de testar o desempenho do aprendiz a partir de uma orientação prévia.

5.2 Apresentação e discussão dos resultados

Nesta subseção, apresentamos os resultados do experimento realizado por Mesquita (2019), com ênfase apenas nos resultados globais – os dados não serão analisados em função dos níveis de proficiência dos participantes, nem em função dos participantes tomados individualmente. Esse recorte tem por objetivo primordial demonstrar o efeito da interferência da L1, por um lado, e desenvolver novas perspectivas de análise dos resultados, em relação ao trabalho citado, por outro.

Inicialmente, retomamos nossa hipótese de trabalho. Com relação à aquisição de L2, adotamos a Hipótese do Acesso Parcial à GU: a L1 é o estado inicial do processo de aquisição de L2, e o aprendiz dispõe apenas dos parâmetros da L1 e dos princípios invariantes da GU (WHITE, 2003). Seguindo White (2003), consideramos a L1 como o estado mental inicial, enfatizando que as representações da interlíngua admitem a interferência da L1 e não exclui o acesso aos princípios invariantes da GU (cf. CRAATS *et al.*, 2000), mediante o acesso ao *input* linguístico. Dessa forma, não é crucial, neste trabalho, investigar o caráter parcial ou total do acesso à GU, mas antes, as propriedades formais da L1 e da língua alvo, esta última definida em função do *input* linguístico disponível ao aprendiz. Se há coincidência paramétrica entre a L1 e a L2, a interferência é positiva, se os valores paramétricos são divergentes, a interferência é negativa. Nesse sentido, a hipótese de trabalho é a de que a interferência positiva da L1 favorece o desenvolvimento linguístico em direção à língua alvo, enquanto a interferência negativa da L1 desfavorece a convergência. As fases da interlíngua consistem nas etapas que conduzem à convergência em relação à língua alvo, mediante o acesso ao *input* da L2 e a interferência positiva, por um lado, e a redução da interferência negativa, por meio de mecanismos de compensação, como alinhamentos por traços semânticos, estratégias de memorização, entre outros, por outro (cf. TSIMPLI; ROUSSOU 1991).

Neste estudo, investigamos a hipótese da interferência da L1, considerando os dados globais gerados pelos participantes, independentemente do grau de proficiência. A análise do resultado global permite verificar, preliminarmente, o efeito da interferência da L1. Passamos, em seguida, à análise dos dados em função do nível de proficiência, que permite verificar os detalhes do desenvolvimento linguístico, ou seja, se o traço examinado é favorecido, em função dos grupos identificados – no caso, iniciante (1), básico (2), intermediário (3). Por hipótese, a interferência positiva da L1 favorece o desenvolvimento linguístico, mediante o acesso ao *input* da L2. Na análise por nível de proficiência, o efeito é examinado pelo ponto de vista do uso da preposição (em oposição à ausência da preposição e demais estratégias).

Com relação ao complemento dativo, partimos da hipótese (cf. seção 4) segundo a qual o PB e a Libras apresentam a mesma estrutura sintática: o complemento dativo é realizado por um núcleo

relacional P/Q(\subseteq), categoria que denota uma relação de transferência de posse entre dois argumentos internos na estrutura de verbos bitransitivos e a uma relação entre um argumento interno e um subevento na estrutura eventiva de verbos monotransitivos (cf. seções 3 e 4) (MANZINI e FRANCO, 2016). No PB, esse núcleo é lexicalizado pela preposição ‘para’ (ou ‘a’, mais restrita à escrita) em verbos bitransitivos e monotransitivos; em Libras, o núcleo P/Q(\subseteq) é lexicalizado pelo morfema direcional (DIR), em verbos de concordância. Em verbos simples bitransitivos (sem concordância), partimos da hipótese de que a transferência é expressa em uma estrutura que inclui a relação entre os dois argumentos internos, Y e Z, como se depreende da estrutura léxico-conceptual (X [CAUSAR Y [TER Z]]), em que o argumento Y é interpretado como alvo e marcado para o traço [+animado], o que explica sua relação, na estrutura do predicado, com o núcleo Q(\subseteq). (cf. 24 acima).

Na análise, realizamos uma comparação entre os grupos de verbos I, II, IIIA, IIIB, IVA e IVB (seção 2), para verificar a interferência da L1 na interlíngua. Na investigação, observamos dois pontos principais: (a) ocorrência de preposições em geral, em oposição a ausência de preposição (SP), e outras estratégias de estruturação do predicado; (b) ocorrência da preposição ‘para’, em relação à presença do núcleo relacional na estrutura do verbo em Libras. Apresentamos, a seguir, os subtotais para as variáveis PREPOSIÇÃO e DEMAIS estratégias (cf. TABELA 1):

Tabela 1: Resultado do teste - uso de preposições e uso de ‘para’

		GRUPO I	GRUPO II	GRUPO IIIA	GRUPO IIIB	GRUPO IVA	GRUPO IVB
PREPOSIÇÃO	A	7,7%	2,1%	1,4%	0,7%	0,7%	-
	A*	0%	0%	1,4%	-	-	-
	PARA	54,9%	57,9%	70,9%	2,8%	-	-
	COM	2,0%	1,0%	0,8%	3,1%	1,0%	0,7%
	DE	1,4%	0,7%	0,2%	0,7%	2,3%	0,2%
	EM	-	0,0%	2,8%	-	8,3%	0,8%
DEMAIS	FOR	-	-	-	0,7%	-	-
	SOBRE	-	-	-	-	0,7%	-
Subtotal PREP		49,7%	46,2%	26,2%	8,3%	35,9%	7,6%
DESAUSENTE	SP	50,0%	3,0%	6,8%	87,0%	53,8%	55,0%
	DIR	0,7%	0,7%	-	-	-	0,7%
	SC	1,4%	1,0%	3,4%	1,1%	4,8%	0%
	SV	1,4%	0,7%	1,0%	0,7%	4,1%	4,8%
	SI	1,0%	2,1%	1,0%	-	0,7%	-
DESAUSENTE	NR	-	-	-	1,4%	0,7%	-
	Subtotal	50,3%	53,8%	73,8%	91,7%	64,1%	92,4%
TOTAL		100%	100%	100%	100%	100%	100%

Fonte: elaboração própria

No cenário ‘PREPOSIÇÃO’, encontramos a porcentagem do uso de cada preposição empregada por grupo de verbos e, ao final (‘Subtotal PREP’), o total das sentenças preposicionadas. Conforme ilustrado, as preposições empregadas foram ‘a’, ‘a*²⁰’, ‘para’, ‘com’, ‘de’, ‘em’, ‘por’ e ‘sobre’. Já no cenário ‘DEMAIS’, encontram-se sentenças sem preposição (SP), sentenças com elemento de outra categoria gramatical (OUTRO), sentenças sem complemento (SC), casos em que o participante não empregou o verbo solicitado (NV), sentenças de difícil interpretação (NI) e casos em que o participante não respondeu a questão (NR). O total geral (TOTAL) é a soma do subtotal dos dois quadros.

Assim, apresentamos inicialmente os resultados, tendo como foco a investigação comparativa entre os grupos de verbos, para verificar a interferência da estrutura da Libras em relação ao uso das preposições na interlíngua. A previsão é que o uso de preposições será favorecido nas estruturas em que o verbo do PB corresponder a um verbo que selecione núcleo de inclusividade, P/Q(⊆), em Libras: verbos de concordância (grupos I, IIIA e IIIB) ou verbos simples bitransitivos que selecionem um argumento alvo, também marcado pelo núcleo de inclusividade (Q(⊆)) (grupo II). Esse uso será convergente ou não com o português, a depender da escolha da preposição, e da presença ou não da preposição na estrutura do predicado na língua alvo. Diferentemente, nas estruturas que correspondem a verbos monotransitivos simples (sem concordância) (grupo IV), o aprendiz deverá desenvolver esse conhecimento com base nos dados do *input* linguístico e não por interferência da Libras (já que o verbo correspondente em Libras não é marcado por DIR nem manifesta o núcleo de inclusividade). Nesse sentido, a previsão é que as estruturas sintáticas similares nas duas línguas favorecem a produção de estruturas convergentes (interferência positiva da L1).

Considerando o subtotal em relação a variável dependente ‘presença de preposição’, observamos, na Tabela 1, que o maior número de complementos preposicionados ocorreu nos grupos I (49,7%) e II (46,2%), em oposição às ‘demais estratégias’ (sem preposição (SP) e outras). De acordo com diversos modelos de análise de aquisição de L2 (AL2), a opcionalidade é um fenômeno que caracteriza a gramática da interlíngua (Sorace, 2003). Nesse sentido, a existência de variação permite extrair conclusões sobre as características da interlíngua e do processo de AL2 quanto ao desenvolvimento linguístico. A comparação com os demais grupos evidencia um percentual menor no uso da preposição, em ordem decrescente, nos grupos IVA (35,9%), IIIA (26,2%) e IIIB (8,3%). A menor taxa de uso da preposição é observada no grupo IVB (7,6%). Em perspectiva global, o uso da

20 Em alguns casos, não foi possível identificar se o elemento ‘a’ empregado se tratava de preposição ou artigo (como, por exemplo, nesta sentença: O Cascão telefonou a Mônica (CM, 3)). Optamos por considerá-la uma preposição, mas, tendo em vista a nossa incerteza, identificamos esse elemento com um asterisco.

preposição na estrutura do predicado como introdutora do argumento interno na interlíngua pode ser considerado evidência de desenvolvimento linguístico, por indicar conhecimento em relação ao uso dessa categoria nesse contexto sintático – ausente na L1 (o movimento direcional (DIR) é considerado um morfema codificador do caso dativo, mas não uma preposição).²¹

Em relação ao uso da preposição ‘para’, os resultados são os seguintes. O grupo I, que inclui verbos bitransitivos (‘dar’, ‘ensinar’, ‘responder’, ‘contar’) e monotransitivos (‘telefonar’), que selecionam a preposição dativa no PB e o movimento direcional (DIR) nos verbos correspondentes em Libras, parece confirmar a hipótese inicial de trabalho de que os verbos de concordância interferem positivamente na aquisição de preposições dativas do português, confirmando-se a relação dessa categoria com o núcleo P/Q(\subseteq), lexicalizado pelo morfema DIR em Libras. Observa-se, em especial, a presença da preposição ‘para’ em 34,5% das ocorrências, contra 15,2% de todas as outras preposições (cf. 25):

(25) Mônica dá comida para o coelhinho. (RC, 2)

No grupo II, de verbos bitransitivos (como ‘explicar’, ‘escrever’, ‘falar’, ‘dizer’, ‘prometer’), que selecionam a preposição dativa no PB, mas não apresentam o morfema DIR (verbo simples) no verbo correspondente em Libras, também observamos um alto número de sentenças preposicionadas (46,2%) (tabela 1). Nesse caso, os surdos também utilizaram muitas preposições, com destaque para a preposição ‘para’ (31,7%), contra 14,5% em relação ao uso de outras preposições (cf. 26):

(26) O Cascão **disse para** a Mônica. “Oi”. (DA, 3)

Como proposto na seção 4, o núcleo P/Q(\subseteq) é selecionado na estrutura dos verbos do grupo II, e introduz um argumento marcado para o traço [+animado], tendo em vista a hipótese de que apresentam um operador de transferência em sua estrutura léxico-conceitual. Enquanto no PB e nos verbos de concordância em Libras, o núcleo de inclusividade é lexicalizado pela preposição ‘para’ e por ‘DIR’, respectivamente, no caso dos verbos simples em Libras, DIR não está lexicalizado como um morfema independente na estrutura do verbo. Diante disso, mediante a comparação com línguas em que essa propriedade está presente no sistema pronominal, como no caso dos pronomes marcados para o caso dativo, (como ‘lhe’, do português, ‘le’, do espanhol), propomos que o núcleo de inclusividade é lexicalizado na estrutura do DP pronominal como uma categoria do tipo Q(\subseteq), por ser marcado para o traço de animacidade. Por hipótese, a presença do operador de transferência

21 Remetemos o leitor ao estudo de Mesquita & Salles (2010) para uma discussão em relação à presença da categoria preposição na Libras, em que se conclui que essa categoria está presente na língua no inventário das categorias lexicais, excluindo-se o contexto da complementação verbal.

na estrutura léxico-conceitual desses predicados determina que o núcleo $Q(\underline{\quad})$ seja projetado na estrutura sintática do DP em Libras, que realiza o argumento alvo. Nesse sentido, como nos verbos do grupo I, com movimento direcional (DIR), pode-se supor que a presença desse núcleo, nos verbos do grupo II, interfere positivamente na aquisição da preposição do português.

O grupo III é dividido no subgrupo IIIA, de verbos monotransitivos (como ‘zombar’, ‘cuidar’, ‘atirar’, ‘mandar’, ‘bater’), que selecionam preposição não dativa no PB, e movimento direcional (DIR) no verbo correspondente em Libras, e no subgrupo IIIB, de verbos monotransitivos (como ‘ajudar’, ‘acusar’, ‘ver’, ‘vencer’, ‘abandonar’), que não selecionam preposição no PB, mas apresentam movimento direcional (DIR) no verbo correspondente na Libras. De acordo com a presente análise, a estrutura sintática dessas construções difere nas duas línguas. Com relação aos verbos do grupo IIIA, por hipótese, existe uma relação de transferência; no entanto, em português, a preposição utilizada não marca orientação para um alvo, mas o ponto em que se situa esse alvo (o que explica o uso das preposições locativas ‘de’ e ‘em’). Nesse sentido, o núcleo $P/Q(\underline{\quad})$ não é selecionado no PB, e o argumento interno é realizado como um complemento oblíquo (preposicionado). Diferentemente, em Libras, os verbos desse grupo são realizados por movimento direcional (DIR). Nesse sentido, o núcleo relacional $P/Q(\underline{\quad})$ está presente nesses verbos, lexicalizado por DIR, uma vez que denota orientação para um alvo. Já no grupo IIIB, no PB, os verbos realizam o argumento interno sem preposição (objeto direto); enquanto em Libras, essas construções também são analisadas como tendo um núcleo $P/Q(\underline{\quad})$, lexicalizado por DIR. Em IIIA e IIIB, DIR denota uma relação entre o argumento interno e o componente nominal da estrutura do predicado. Conforme mencionado anteriormente, a relação com o componente nominal da estrutura do predicado pode ou não estar codificada por meio de um núcleo relacional – o que explica o contraste translinguístico entre o PB e a Libras (cf. seção 3 e 4).

Em IIIA, observamos um alto número de estruturas não preposicionadas (64,8%), superando as construções com preposição (26,2%), apesar da presença do morfema DIR em Libras. Esse alto índice de estruturas sem preposição pode ser explicado por uma divergência entre a estrutura da L1 e os dados do *input*: em Libras, a marcação do argumento interno se dá pelo morfema DIR, que realiza um movimento em direção ao alvo; já no PB, como mencionado, as preposições denotam o ponto (estático) de localização do argumento interno, sendo, portanto, um tipo de complemento locativo. Assim, a preposição no PB não denota direcionalidade, embora as propriedades léxico-conceituais do verbo nas duas línguas descrevam uma orientação para um alvo. Deste modo, o alto índice de estruturas sem preposição pode ser explicado diante da inconsistência sintático-semântica entre a estrutura da L1 e o uso da preposição nos dados do *input*. Entretanto, é interessante observar que, nas estruturas em que ocorre o uso da preposição, a preposição ‘para’ (divergente do PB) é a mais

empregada, com 7,6% de ocorrência. Conclui-se que existe interferência (positiva) da L1 em relação ao uso de preposição nesse contexto sintático, determinada pela presença do morfema DIR nos verbos correspondentes da Libras. No entanto, em relação à escolha do item lexical, não sendo convergente com a língua alvo, caracteriza a interferência (negativa) da L1 (cf. 27):

(27) O Cebolinha e Cascão se **zombaram** para a Mônica. (DA, 3)

Outra preposição que ocorre no grupo IIIA é ‘com’, também divergente em relação à língua alvo. Sua ocorrência nos dados pode estar relacionada ao fato de que essa preposição denota inclusividade. Nesse sentido, a presença da preposição ‘com’ pode ser considerada um efeito de interferência (positiva) da L1, em relação ao uso da preposição nesse contexto sintático. Ainda que a frequência no uso desse item seja baixa, sua ocorrência, considerada juntamente com o uso da preposição ‘para’, aponta para a tendência de favorecimento na presença do núcleo de inclusividade P/Q(\subseteq) nos verbos desse grupo em Libras (cf. 28), embora não convergente, em relação à escolha do item lexical:

(28) A Mônica **bate** com Cebolinha. (GC, 2)

Passemos, agora, ao grupo IIIB, que também apresenta mais sentenças com o argumento interno sem preposição (87,6%) do que preposicionadas (8,3%) (cf. tabela 1), mesmo com a presença do morfema DIR em Libras. No PB, o argumento interno dos verbos desse grupo é realizado sem preposição. Uma explicação para esse resultado seria a ausência da preposição nos dados do *input*. No entanto, assim como em IIIA, merece referência o uso de preposição na estrutura do predicado e, particularmente, o índice mais alto da preposição ‘para’, o que remete à estrutura dos verbos desse grupo em Libras, em que se verifica a presença de DIR. Ainda que a frequência seja muito baixa, como evidência de opcionalidade, essas ocorrências podem ser analisadas como interferência (negativa) do núcleo relacional (DIR) da Libras (cf. 29) (cf. para um resultado semelhante, veja-se Mesquita & Salles, 2010):

(29) A Magali **acusou** para o gato. (CB, 3)

Os resultados em relação aos grupos I, II, IIIA são consistentes com a hipótese de trabalho, no sentido de demonstrar a interferência do morfema DIR, na estrutura dos verbos da L1, em relação ao uso da preposição ‘para’: ainda que com frequências diferentes, e consequência diferentes para a convergência, pelas razões referidas anteriormente, verifica-se que o uso da preposição ‘para’ se destaca em relação às demais preposições. O grupo IIIB só não se alinha com os anteriores pelo fato de que o uso da preposição ‘com’ concorre com o de ‘para’, mas a manifestação de ambas pode ser vinculada ao morfema DIR, pois denotam relação de inclusividade.

O grupo IV é dividido no subgrupo IVA, formado por verbos monotransitivos (como ‘pensar’, ‘gostar’, ‘precisar’, ‘concordar’, ‘confiar’), que selecionam preposição não dativa no PB e verbos simples na Libras, e no subgrupo IVB, formado por verbos monotransitivos (como ‘amar’, ‘comer’, ‘beber’, ‘ter’, ‘saber’), que selecionam complemento sem preposição (objeto direto) no PB e verbos simples na Libras. Iniciemos por IVA: verificamos que as sentenças com complementos sem preposição ocorrem com maior frequência (53,8%) do que as sentenças com complementos preposicionados (35,9%). Do ponto de vista global, esse resultado está de acordo com a hipótese de trabalho, uma vez que, segundo a presente análise, os verbos simples monotransitivos da Libras não selecionam núcleo $Q(\subseteq)$, que, por hipótese, interfere (positivamente) na aquisição das preposições, particularmente a preposição ‘para’ (cf. 30):

(30) A Magali comida **pensar** _ (SK, 1)

No entanto, merece destaque o percentual de 35,9% de complementos preposicionados, uma vez que a presença de complementos preposicionados (não dativos) no PB configura um contraste paramétrico em relação à Libras. Tendo em vista a hipótese do acesso parcial à GU, assumimos que, na aquisição dessas estruturas, o aprendiz faz uso de estratégias de compensação, que são estratégias independentes da GU. Em particular, o desenvolvimento das estruturas da língua alvo será apoiado na relação composicionalidade semântica entre o verbo e a preposição, mediante o acesso ao *input* da língua alvo – por exemplo, o verbo ‘pensar’ seleciona um argumento tema, que é introduzido na estrutura do predicado por meio da preposição locativa ‘em’. A mesma hipótese se aplica aos outros verbos, em que a preposição selecionada pode ser ‘com’ ou ‘de’. É interessante notar que não há nenhuma ocorrência da preposição ‘para’ nos dados da interlíngua no grupo IVA – o que pode ser uma confirmação indireta da interferência de DIR na aquisição dessa preposição.

O grupo IVB também apresenta uma porcentagem superior de estruturas não preposicionadas em relação às preposicionadas (77,9% e 7,6%, respectivamente) (cf. tabela 1). Uma vez que esse grupo é formado por verbos simples em Libras e por complementos não preposicionados no PB, era esperado que os participantes produzissem mais estruturas sem preposição. A presença das preposições, entretanto, merece consideração: por um lado, não se verifica o uso da preposição ‘para’, o que pode ser uma confirmação (indireta) da hipótese sobre a relação dessa preposição com o morfema DIR (e sua realização sintática como núcleo de inclusividade $P/Q(\subseteq)$); por outro lado, o uso de preposições como ‘com’, ‘de’ e ‘em’, ainda que com baixa frequência, aponta para a situação de opcionalidade, típica do desenvolvimento linguístico de L2. Apesar da coincidência paramétrica entre o PB e a Libras, no sentido de realizar o complemento como um objeto direto (supostamente, a opção

default), é possível supor que o aprendiz generaliza para esse grupo a estratégia adotada nos outros grupos, em relação ao uso da preposição como introdutora do complemento verbal. Nesse sentido, é possível analisar o uso dessas preposições em termos dos traços lexicais envolvidos – trata-se de preposições cuja conteúdo semântico é compatível com a estrutura léxico-conceitual dos verbos em questão. Por exemplo, com o verbo ‘amar’, em que ocorreu a preposição ‘com’, é possível supor a composicionalidade semântica entre a leitura estativo-comitativa da preposição ‘com’ e a denotação do verbo psicológico ‘amar’ (cf. 31):

(31) O Chico Bento **ama** com a Rosinha. (DA, 3)

Conforme mencionado, os verbos dos grupos IVA e IVB não envolvem a presença do núcleo relacional no português brasileiro e na Libras (verbos simples). Nesse sentido, o uso da preposição introdutora do complemento na estrutura dos verbos monotransitivos do grupo IVA no português (e em outras línguas) não corresponde à marcação do caso dativo. O uso da preposição com esses verbos tem sido analisado como uma propriedade idiossincrática, determinada no nível do léxico (cf. JACKENDOFF, 1990a). Nessa estrutura, o uso da preposição determina uma relação sintática oblíqua, sem haver, porém, um alinhamento formal com o núcleo relacional. A hipótese é que, nesse caso, a distribuição da preposição (e a variação translinguística em relação a sua ocorrência) é uma opção marcada – inversamente, a ausência da preposição é a opção *default*, considerada um princípio da Gramática Universal.

Por sua vez, o uso da preposição na interlíngua é explicado em termos da relação composicional entre as propriedades semânticas do verbo e da preposição. No grupo IVA, essa estratégia resulta em convergência em relação ao uso, mas não necessariamente em relação à escolha, da preposição. Por hipótese, essa estratégia é generalizada para os verbos do grupo IVB, apesar da coincidência paramétrica, em relação à ocorrência do objeto direto no PB e na Libras. Diante do contraste paramétrico entre a L1 e a L2, uma vez que não existe preposição nesse contexto sintático na L1, o uso da preposição com verbos dos grupos IVA e IVB é considerado resultante de uma estratégia de compensação (no caso, a marcação da fronteira sintática com base na composicionalidade semântica).

A análise dos dados mostrou que o uso (em oposição à ausência) da preposição na interlíngua pode ser explicado por interferência (positiva), nos casos em que há coincidência na marcação paramétrica entre a L1 e a L2 em relação à presença do núcleo relacional P/Q(\subseteq) lexicalizado por movimento direcional (DIR) na Libras e por preposição dativa no PB, ou por interferência negativa, em relação ao uso de preposição na presença do movimento direcional (DIR) no verbo correspondente

na Libras e de ausência de preposição no PB. Essa situação é verificada ainda em relação à escolha da preposição – por hipótese, o uso da preposição ‘para’ caracteriza interferência positiva com verbos dos grupos I e II, e negativa, com verbos dos grupos IIIA e IIIB, em que se verifica uso de preposição diferente de ‘para’ ou ausência de preposição, respectivamente.

Na etapa seguinte do estudo, aplicamos o teste estatístico não paramétrico de Fisher (SIEGEL; CASTELLAN JR., 2006) aos resultados obtidos nos grupos I, II e IIIA e IIIB. Considerando-se a ausência categórica de ‘para’ nos dados referentes aos grupos IVA e IVB, e por corresponderem a verbos simples na Libras, estando ausente o núcleo relacional P/Q(\subseteq), esses grupos foram retirados da análise estatística. Assumindo-se que o efeito da interferência do núcleo relacional P/Q(\subseteq) da L1 se manifesta pela escolha da preposição ‘para’, a hipótese de trabalho é a de que a preposição ‘para’ ocorra com os verbos correspondentes na L2. Considerando-se que o uso da preposição ‘para’ com os verbos do grupo I e II é convergente com a língua alvo, configurando-se interferência positiva, e com os verbos dos grupos IIIA e IIIB é divergente com a língua alvo, configurando-se interferência negativa, a análise considerou os grupos assim amalgamados: I+II e IIIA+IIIB. O resultado do teste quanto ao uso de ‘para’ nos dois grupos é 0,0111, o que corresponde a um p-valor inferior a 0.05. O valor igual ou inferior $p < 0.05$ indica que existe dependência estatística entre os grupos examinados, ou seja, existe concordância entre os grupos em relação às respostas ao ‘uso de para’. Esse resultado rejeita a hipótese nula, e confirma hipótese de trabalho de que existe semelhança entre os grupos em relação ao favorecimento de ‘para’ na presença do traço relevante (com efeito positivo e negativo, respectivamente, quanto à convergência com a língua alvo).

Apresentamos, a seguir, os dados percentuais relativos ao uso da preposição, em oposição à ausência da preposição e demais estratégias, em função dos níveis de proficiência, distribuídos nos seguintes níveis: nível 1: iniciante; nível 2: básico; nível 3: intermediário (cf. 5.1.1). Observa-se a tabela 2 abaixo:

Tabela 2: Resultado: uso de preposições por nível de proficiência

	Nível 1	Nível 2	Nível 3
GRUPO I	0	44%	91%
GRUPO II	0	54%	72,7%
GRUPO IIIA	0	34%	38,2%
GRUPO IIIB	0	8%	14,6%
GRUPO IVA	5%	38%	56,4%
GRUPO IVB	0	14%	7,3%

Fonte: elaboração própria

Considerando-se que a preposição ocorre, na língua alvo, com os verbos dos grupos I, II, IIIA

e IVA, a ocorrência dessa categoria (independentemente da convergência em relação à escolha da preposição) permite verificar o desenvolvimento linguístico. Nos grupos I e II, verifica-se o aumento gradual e significativo no percentual de uso da preposição, em função dos níveis de proficiência (0% > 44% > 91% e 0% > 54% > 72,7%, respectivamente), confirmando-se a hipótese da interferência positiva da L1, uma vez que os verbos nesses grupos manifestam DIR.

No grupo IIIA, em que os verbos na L1 manifestam DIR, verifica-se aumento no percentual de uso da preposição, em função dos níveis de proficiência (0% > 34% > 38,2%). Embora o índice seja menor se comparado ao dos grupos I e II, é possível afirmar que existe desenvolvimento linguístico, uma vez que os complementos desses verbos ocorrem com preposição, na língua alvo, confirmando-se a hipótese de interferência positiva da L1 (independentemente da convergência em relação à escolha da preposição). Em relação ao grupo IIIB, em que os predicados da língua alvo não apresentam preposição, verifica-se o aumento gradual do percentual de uso da preposição, em função do nível de proficiência, ainda que os índices não sejam altos (0% > 8% > 14,6%). O uso da preposição, nesse caso, pode indicar uma situação de interferência negativa da L1, uma vez que os verbos desse grupo manifestam DIR.

Inversamente, o alto índice da estratégia de não uso da preposição, opção convergente com a língua alvo, pode ser atribuído ao caráter não marcado dessa estrutura (V+OD), disponível na GU como uma opção *default*.

Nos verbos do grupo IVA, que não manifestam DIR, tem-se aumento no uso da preposição, em função do nível de proficiência (5% > 38% > 56,4%), o que configura desenvolvimento linguístico. Conforme mencionado anteriormente, o uso da preposição na estrutura desses predicados configura uma opção marcada, que a gramática da L1 não adota. Por hipótese, o desenvolvimento linguístico, nesse caso, deve estar apoiado em alguma estratégia de compensação (uma questão que deixamos em aberto neste estudo). Finalmente, nos verbos do grupo IVB, tem-se um resultado particularmente interessante, pois o percentual de uso da preposição tem um aumento do nível 1 para o nível 2, e depois um declínio (0% > 14% > 7,3%). Considerando que os predicados desse grupo não apresentam preposição na língua alvo, conclui-se que o declínio configura desenvolvimento linguístico, evidenciando-se a tendência em direção à convergência com a língua alvo.

Considerações finais

O presente artigo investigou a ocorrência de preposições dativas na interlíngua do surdo, considerando a interferência dos verbos de concordância da Libras. Tendo em vista as propriedades

morfofossintáticas desses verbos na codificação do argumento alvo, as características da interlíngua permitem concluir que a interferência da L1 se manifesta pelo favorecimento no uso das preposições, em oposição ao não uso e às demais estratégias, nos verbos correspondentes da L2. Em particular, os resultados demonstram que a estrutura dos verbos de concordância em Libras influencia na aquisição da preposição em complementos dativos no português, o que se estende para outros contextos preposicionados. Considerando a ausência de preposição em contexto de complementação verbal na Libras, o uso de preposição, na interlíngua, é considerado desenvolvimento linguístico. Nesse sentido, o favorecimento da preposição ‘para’, associado aos casos em que o verbo correspondente da L1 é de concordância, é considerado evidência para a hipótese da interferência da L1.

Inicialmente, analisamos as propriedades do complemento dativo nas LOs. Assumimos que esse complemento é licenciado pelo núcleo Q/P(\subseteq), tanto na estrutura de predicados bitransitivos quanto na estrutura de predicados monotransitivos (MANZINI; FRANCO, 2016). Adotamos essa mesma análise para estruturas com verbos de concordância em libras, propondo que o argumento dativo é introduzido por Q/P(\subseteq), que é realizado lexicalmente pela categoria DIR, tanto em predicados bitransitivos quanto em predicados monotransitivos. Por hipótese, esse núcleo também está presente na estrutura de verbos simples bitransitivos na Libras, como um traço na estrutura do DP pronominal marcado para o traço [+animado].

Apresentamos, em seguida, uma análise dos dados distribuído nos seguintes grupos de verbos: I – dativo no PB/concordância na libras; II – dativo no PB/simples (bitransitivo) na libras; IIIA – não dativo preposicionado no PB/ concordância na libras; IIIB – não dativo sem preposição no PB/ concordância na libras; IVA – não dativo preposicionado no PB/simples (monotransitivo) na Libras; IVB – não dativo sem preposição no PB/simples (monotransitivo) na Libras. Discutimos a relação entre o núcleo relacional Q/P(\subseteq), presente em verbos de concordância (grupos I, IIIA, IIIB) e em verbos simples bitransitivos (grupo II), e o uso da preposição (em oposição à ausência e ao uso de outras estratégias) na interlíngua. Mereceu destaque a ausência da preposições ‘para’ na interlíngua nos dados do grupo IV, em que o núcleo relacional Q/P(\subseteq) está ausente nos verbos da L1 (por serem verbos simples monotransitivos).

Apresentamos, em seguida, o resultado do teste estatístico (Fisher) realizado com o objetivo de investigar a significância dos resultados obtidos nos grupos I, II, IIIA e IIIB. Verificou-se a dependência estatística entre os grupos amalgamados I-II, em oposição a IIIA-IIIB em relação às respostas dos participantes ao uso de ‘para’. A dependência estatística nos resultados permitiu concluir que existe semelhança entre os grupos no efeito da interferência do traço relacional Q/P(\subseteq) dos verbos de

concordância em relação ao uso da preposição ‘para’ na interlíngua.

Finalmente, apresentamos a análise dos dados em função da proficiência dos participantes, distribuídos nos níveis iniciante (1), básico (2) e intermediário (3). Demonstrou-se que existe desenvolvimento linguístico para o uso da preposição, em oposição à ausência da preposição e demais estratégias, na presença do núcleo relacional Q/P(⊆) na estrutura do predicado. Outros aspectos devem ser considerados como, por exemplo, o papel do *input* no desenvolvimento linguístico e as implicações da escolha da preposição. Essas questões, de suma relevância, serão consideradas em trabalhos futuros.

REFERÊNCIAS

- ANAGNOSTOPOULOU, Elena. *The Syntax of Ditransitives. Evidence from Clitics*. Berlin-New York: Mouton de Gruyter, 2003.
- BAKER, Mark. *Incorporation*. Chicago: University of Chicago Press, 1988.
- BARSS, Arnold; Howard LASNIK. A Note on Anaphora and Double Objects. *Linguistic Inquiry* 17: 347-354, 1986.
- BECHARA, Evanildo. *Moderna Gramática Portuguesa*. 42ª edição. Rio de Janeiro: Lucerna, 2004.
- BENVENISTE, Émile. *Problemas de Linguística Geral*. Campinas: Pontes, (1ª edição 1966), 1988.
- BERLINCK, Rosane Andrade. The Portuguese dative. In William Van Belle and Willy Van Langendonck (eds) *The Dative. Descriptive Studies*, 119-151. Amsterdam: John Benjamins, 1996.
- BORER, Hagit. ‘Passives without Theta Grids’. Lapointe (ed) *Morphological Interfaces*, CSLI, Stanford, 1996.
- CAPOVILLA, Raphael et al. *NOVO DEIT-LIBRAS: Dicionário Enciclopédico Ilustrado Trilíngue da Língua de Sinais Brasileira*. Vol. 1. 2. São Paulo: EDUSP, 2012.
- CHOMSKY, Noam. *O Conhecimento da Língua – sua natureza, origem e uso*. Trad. Anabela Gonçalves e Ana Teresa Alves (1994). Lisboa: Caminho, 1986.
- _____. *O Programa Minimalista*. Trad. Eduardo Paiva Raposo. Cambridge: MIT Press, 1995.

COLORIR.ORG. Disponível em <https://colorir.org>. Acesso em 10 de janeiro de 2020.

CRAATS, Ineke van de; CORVER, Norvin.; HOUT, Roeland van. Conservation of grammatical knowledge: on the acquisition of possessive noun phrases by Turkish and Moroccan learners of Dutch. *Linguistics* 38-2, 221-314, 2000.

CUERVO, Maria Cristina. Datives at Large. PhD thesis. Massachusetts Institute of Technology, 2003.

DEMONTE, Violeta. Dative Alternation in Spanish. *Probus* 7: 5-30, 1995.

EMONDS, Joseph. Projecting indirect objects. *The Linguistic Review* 10, 211-263, 1993.

FELIPE, Tanya; MONTEIRO, Myrna. *Libras em contexto – curso básico* (livro do professor). Brasília: Ministério da Educação, Secretaria de Educação Especial, 2007.

FERNANDES, Eulália. *Linguagem e surdez*. Porto Alegre: Artes Médicas, 2003.

FERREIRA-BRITO, Lucinda. *Por uma gramática da língua de sinais*. Rio de Janeiro: Tempo Brasileiro, 1995.

HALE, Ken.; KEYSER, Samuel Jay. On argument structure and the lexical expression of syntactic relations. In HALE, Ken.; KEYSER, Samuel Jay (ed.) *The view from the building 20*. Cambridge, Mass.: The MIT Press, 1993. p. 53–109.

HARLEY, Heidi. Possession and the Double Object Constructions. In PICA, Pierre and Johann ROORYCK. *Linguistic Variation Yearbook 2*, 31-70. Amsterdam: John Benjamins, 2003.

JANIS, Wynne. A Cross-linguistic Perspective on ASL Verb Agreement. In: EMMOREY, Karen.; REILLY, Judy (eds.) *Language, Gesture, and Space*. Hillsdale, NJ: Lawrence Erlbaum, 1995. p. 195-223.

JACKENDOFF, Ray. *Semantic Structure*. Cambridge, Mass.: MIT Press, 1990a.

_____. On Larson's treatment of the double object construction. *Linguistic Inquiry*, 21 (3), 1990b, 27-456.

KAYNE, Richard. *Connectedness and Binary Branching*. Dordrecht: Foris, 1984.

LARSON, Richard. On the double object construction. *Linguistic Inquiry* 19 (3), 335-91, 1988.

LEVIN, Beth; HOVAV, Malka Rappaport. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, Mass.: MIT Press, 1995.

LENNEBERG, Eric. *Biological Foundations of Language*. New York: J. Wiley, 1967.

MANZINI, Maria Rita.; FRANCO, Ludovico. Goal and DOM datives. *Natural Language and Linguistic Theory* 34 (1), 2016. p. 197-240

MARANTZ, Alec. *On the Nature of Grammatical Relations*. MIT Press, Cambridge, Mass, 1984.

_____. Implications of asymmetries in double object constructions. In Sam MCHOMBO (ed.) *Theoretical Aspects of Bantu Grammar*, Stanford, CA: CSLI Publications, 1993, p. 17–46.

MEIR, Irit. A Cross-Modality Perspective on Verb Agreement. *Natural language and Linguistic Theory* 20: 413-450, Kluwer Academic Publishers. Printed in the Netherlands, 2002.

MESQUITA, Aline. *A categoria preposicional na interlíngua do surdo aprendiz de português (L2)*. 2008, 106 f. Dissertação (Mestrado em Linguística). Universidade de Brasília, Brasília, 2008.

_____. *Estruturas dativas do português (L2) na interlíngua de surdos*. 2019, 281 f. Tese (Doutorado em Linguística). Universidade de Brasília, Brasília, 2019.

MESQUITA, Aline; SALLES, Heloisa M. Lima. Preposições na língua de sinais brasileira e na interlíngua de surdos aprendizes de português L2. In SALLES, Heloisa M. Lima; NAVES, Rozana R. *Estudos gerativos da Língua de Sinais Brasileira e de aquisição de português (L2) por surdos*. Cànone, Goiânia, 2010, p. 157-188.

PADDEN, Carol. *Interaction of Morphology and Syntax in American Sign Language*. PhD Dissertation, University of California, San Diego, 1983 [Publicado em 1988 em *Garland Outstanding Dissertations in Linguistics*, New York]

PERLMUTTER, David. ‘Predicate’: A Grammatical Relation. *Linguistics Notes from La Jolla* 6, 1979.

PESETSKY, David. *Zero Syntax*. Cambridge, Mass: MIT Press, 1995.

PYLKKÄNEN, Lina. *Introducing arguments*. Cambridge, Mass.: MIT PhD Thesis, 2002.

QUADROS, Ronice. M. *Phrase structure of Brazilian Sign Language*. Tese de Doutorado. PUCRS.

Porto Alegre, 1999

QUADROS, Ronice M.; KARNOPP, Lodenir. B. *Língua Brasileira de Sinais: Estudos Linguísticos*. Porto Alegre: Artmed, 2004.

QUADROS, Ronice M.; QUER, Josep. Revertendo os verbos reversos e seguindo em frente: sobre concordância, auxiliares e classes verbais em línguas de sinais. In: Ronice QUADROS; Maria L. VASCONCELLOS. *Questões Teóricas das Pesquisas em Línguas de Sinais*. Petrópolis: Editora Arara Azul, 2008

_____. A caracterização da concordância nas línguas de sinais. In: SALLES, Heloisa M. Lima; NAVES, Rozana R. (orgs.) *Estudos gerativos da Língua de Sinais Brasileira e de aquisição de Português (L2) por surdos*. Goiânia: Cãnone Editorial, 2010, p. 33-58.

RAMOS, Jania. *Marcação sintática e mudança sintática no português*. Campinas, SP: Universidade Estadual de Campinas, Tese de Doutorado, 1992.

RAMCHAND, Gillian. Catriona. *Verb meaning and the lexicon: a first phase syntax*. New York: Cambridge University Press, 2008.

RATHMANN, Christian; MATHUR, Gaurav. Is Verb Agreement the Same Crossmodally? In: MEIER, Richard/Cormier, KEARSY/Quinto-Pozos, David (eds.), *Modality and Structure in Signed and Spoken Languages*. Cambridge: Cambridge University Press, 2002, p. 370-404.

SALLES, Heloisa M. Lima. *Prepositions and the syntax of complementation*. PhD Dissertation. Bangor, 1997, 157 p.

SALLES, Heloisa M. Lima *et al.* “Enunciados inferenciais e estrutura gramatical na interlíngua de surdos aprendizes de português L2”. *Revista Espaço* n 44, 105-124. Rio de Janeiro: INES, 2015.

SIEGEL, Sidney; John CASTELLAN JR. *Estatística não-paramétrica para ciências do comportamento*. Tradução Sara Correa Carmona. Porto Alegre, Artmed, [1988] 2006.

SKLIAR, Carlos. *A surdez: um olhar sobre as diferenças*. Porto Alegre: Mediação, 1998.

SORACE, Antonella. Near-nativeness. In DOUGHTY, C. e LONG, M. (eds.) *Handbook of Second Language Acquisition*. Oxford: Blackwell, 2003.

SVENONIUS, Peter. Icelandic case and the structure of events. *Journal of Comparative Germanic*

Linguistics 5 (1-3), 197-225, 2002.

TORRES MORAIS, Maria Aparecida; BERLINCK, Rosane. A caracterização objeto indireto no português: aspectos sincrônicos e diacrônicos”. In LOBO, Tania et al. (eds.) *Para a história do português brasileiro: Novos dados, novas análises*. Vol. VI. Tomo I, 73–106. Salvador, BA: Editora da Universidade Federal da Bahia, 2007.

TORRES MORAIS, Maria Aparecida.; SALLES, Heloisa M. Lima. Parametric change in the grammatical encoding of indirect objects in Brazilian Portuguese. *Probus*. n. 22, 181–209, 2010.

TSIMPLI, Ianthi Maria; ROUSSOU, Anna. Parameter Resetting in L2? *UCL Working Papers in Linguistics* 3: 149-169, 1991.

WHITE, Lydia. *Second Language Acquisition and Universal Grammar*. Cambridge University Press, Cambridge, 2003.

O IMPACTO DO USO LINGUÍSTICO ORAL EM L2 NO RECONHECIMENTO DE PALAVRAS POR BILÍNGUES TARDIOS DO PORTUGUÊS E DO INGLÊS *THE IMPACT OF L2 ORAL LANGUAGE USAGE ON WORD RECOGNITION BY PORTUGUESE AND ENGLISH LATE BILINGUALS*

Ricardo Augusto de Souza¹
Jamila Viegas Rodrigues²

RESUMO

O objetivo desta pesquisa foi investigar a hipótese de modulação do uso de L2 sobre o reconhecimento por falantes bilíngues do português brasileiro e do inglês de palavras escritas em inglês após sua apresentação oral. Os pares de palavras alvo se distinguem pelas consoantes nasais /m/ e /n/ em final de sílaba (ex.: *them* /ðem/ ‘eles, elas’; *then* /ðen/ ‘então’) porque o contraste fonotático na organização silábica do inglês e do português acarreta impactos sobre o valor fônico atribuído por leitores a grafemas compartilhados pelas convenções ortográficas dessas línguas e que ocorrem em posições grafotáticas análogas na ortografia de sílabas das duas línguas. A investigação ocorreu a partir do rastreamento ocular em experimentos baseados no Paradigma Mundo-Visual (TANENHAUS *et al.*, 1995; TANENHAUS; SPIVEY-KNOWLTON, 1996). O construto “uso linguístico oral” foi operacionalizado através de auto-declaração dos participantes em resposta a itens adaptados do questionário utilizado por Valadares (2017). Os participantes foram 28 brasileiros e nove falantes nativos de língua inglesa. A análise mostrou que maior uso linguístico oral em L2 afeta a capacidade discriminatória dos falantes não nativos. Nossos resultados indicam que o uso é um fator tão relevante na constituição de representações sonoras do léxico em L2 quanto se hipotetiza, a partir da Teoria de Exemplares (BYBEE, 2001, 2010; JOHNSON, 1997, 2005; PIERREHUMBERT, 2001, 2003), que o é na constituição de representações análogas em L1.

Palavras-chave: Bilinguismo. Léxico Mental. Lexema. Psicolinguística.

ABSTRACT

The objective of this study was to investigate the hypothesis that L2 usage modulates Brazilian Portuguese and English bilinguals’ recognition of written English words that were first presented orally. The target word pairs are distinguished by the nasal consonants /m/ and /n/ at the end of a syllable (eg: *them* /ðem/; *then* /ðen/). Although the Roman alphabet letters that usually represent these nasal consonants occur in similar syllable graphotactic positions in both languages, there is a phonotactic contrast between English and Portuguese syllabic organization that has an impact on the phonic value attributed to graphemes shared by the spelling rules of both languages. The investigation was based on the Visual-World Paradigm (TANENHAUS *et al.*, 1995; TANENHAUS; SPIVEY-KNOWLTON, 1996) and its eye-tracking related experiments. Oral language usage was operationalized as participants’ self-declared behaviors collected through items of questionnaire adapted from Valadares (2017). The participants were 28 Brazilians and a control group of nine native English speakers. Our results showed an effect of the amount of L2 oral language use on the discriminatory ability of non-native speakers of English. We discuss the role of the usage patterns captured by the questionnaire items in the acquisition of L2 phonological representations, and the connections between our findings and predictions in Exemplar Theory (BYBEE, 2001, 2010; JOHNSON, 1997, 2005; PIERREHUMBERT, 2001, 2003) for analogous representations in the L1.

Keywords: Bilingualism. Mental Lexicon. Lexeme. Psycholinguistics.

1 Universidade Federal de Minas Gerais/CNPq; email: ricsouza@ufmg.br

2 Universidade Federal de Lavras/DEL; email: jamila.rodrigues@ufla.br

Introdução

Este artigo relata a investigação do acesso lexical no âmbito lexêmico e a sua relação com os perfis de uso linguístico em segunda língua-L2 à luz da Teoria de Exemplos (JOHNSON, 1997; PIERREHUMBERT, 2003) como um modelo baseado no uso (BYBEE, 2001, 2010). O destaque para o âmbito lexêmico do acesso lexical vem da divisão da entrada lexical em dois níveis na arquitetura mental, o **lema** que envolve informações sintáticas e semânticas, e o **lexema** que envolve informações morfológicas e fonológicas (LEVELT, 1989; LEVELT *et al.*, 1999). Primeiramente, apresentamos o conceito de acesso lexical no âmbito lexêmico adotado. O acesso lexical consiste em buscar e recuperar uma entrada lexical, incluindo seu significado, classe gramatical, estrutura sintática e representação fonêmica dentro do léxico mental (FERNANDEZ; CAIRNS, 2010, p. 141). O léxico mental é o repertório em que estão armazenadas as informações sobre as palavras que o falante conhece. De acordo com Guan *et al.* (2019, p. 455), o acesso lexical é também conhecido como reconhecimento de palavra e, neste artigo, serão utilizados como sinônimos.

Embora o resultado do acesso lexical no âmbito lexêmico seja eficiente no reconhecimento de palavras que estão sendo faladas em um ambiente sem muito barulho para pessoas que falem a língua utilizada, a investigação dos processos subjacentes a esse acesso é complexa. Ou seja, a investigação do que ocorre na mente do falante enquanto palavras são acessadas é obscura porque os modelos de acesso lexical fazem afirmações distintas sobre a natureza da organização do léxico mental e sobre o quão tolerante o sistema de processamento é com as incompatibilidades fonológicas no discurso (ALLOPENA; MAGNUSON; TANENHAUS, 1998).

Após apresentarmos o conceito de acesso lexical no âmbito lexêmico e antes de mostrar como foi feita a investigação, é preciso expor nosso objetivo e nossa hipótese principal. O objetivo geral desta pesquisa foi investigar o acesso lexêmico por bilíngues a partir de palavras que se distinguem pela consoante nasal /m/ e /n/ em final de sílaba em inglês por falantes do Português Brasileiro-L1 em relação ao perfil de uso linguístico oral em L2 à luz da Teoria de Exemplos (JOHNSON, 1997; PIERREHUMBERT, 2003) como um modelo baseado no uso (BYBEE, 2001, 2010). A hipótese principal é que a intensidade de competição para o acesso lexêmico seja modulada majoritariamente pelo perfil de uso linguístico oral em L2.

A investigação da competição para o acesso lexêmico ocorreu a partir do rastreamento ocular em experimentos baseados no paradigma mundo-visual (TANENHAUS *et al.*, 1995; TANENHAUS;

3 A distinção entre lema e lexema foi adotada por Levelt (1989, p. 187), mas tal terminologia foi discutida em Kempen e Hoenkamp (1987, p. 214) e mencionada em Kempen e Huijbers (1983, p. 197).

SPIVEY-KNOWLTON, 1996), em que o movimento ocular é mapeado durante exposição simultânea ao estímulo auditivo e à apresentação de quatro imagens ou palavras em tela. A literatura mostra que a sobreposição das duas modalidades (oral e escrita) faz com que a atenção (fixação ocular) seja direcionada para o referente visual e este direcionamento da atenção indica o processamento da linguagem (HUETTIG; MCQUEEN, 2007; MCQUEEN; VIEBAHN, 2007; WEBER *et al.*, 2007).

Allopena *et al.* (1998, p. 438) e Tanenhaus *et al.* (2000, p. 558) apontam que o paradigma mundo-visual é uma metodologia valiosa para o estudo do acesso lexical no âmbito lexêmico por causa da sensibilidade temporal das medidas de fixação ocular. O rastreamento ocular representa, portanto, um grande avanço na habilidade de medir microestruturas do processamento da linguagem porque permite visualizar a trajetória e não somente fornece o tempo final de reação.

Os dados de rastreamento ocular foram relacionados com o perfil de uso linguístico oral em L2. Os perfis de uso linguístico oral em inglês-L2 dos participantes deste estudo foram quantificados a partir de itens extraídos de um questionário utilizado por Valadares (2017) que investiga a possibilidade de alterações de dominância linguística de brasileiros residentes no Brasil e aprendizes de inglês-L2, em certas práticas de letramento, especialmente em relação a mídias digitais. Segundo Heredia (1997, p. 38), dominância é a frequência de uso da língua em comparação com outra língua em consonância com a perspectiva de domínios específicos, tal como sugere o Princípio da Complementaridade de Grosjean (2016, p. 83).

Ao delinear os perfis de uso linguístico oral em inglês L2, atendemos uma demanda de detalhamento e aprofundamento do construto proficiência e da variável imersão (residência no país de origem da língua, sendo ela materna ou adicional) que são constantemente evocados na literatura com especificidades pouco uniformes e sem relação com parâmetros de uso linguístico e dominância suficientemente explicitados ou sequer controlados (HULSTIJN, 2012). Por exemplo, alguns pesquisadores utilizam testes reconhecidos de proficiência de inglês como língua estrangeira (VIDAL, 2011) enquanto outros caracterizam os participantes com relação à proficiência em inglês-L2, mas não explicitam como obtiveram a medida.

Com relação à imersão em país de segunda língua, alguns autores (BUNDGAARD *et al.*, 2011, 2011b; MUNRO; DERWING, 2008) analisam o desenvolvimento perceptual dos participantes bilíngues tardios de acordo com o tempo de residência no país de segunda língua e afirmam que as mudanças significativas ocorrem entre seis e doze meses de imersão. Porém, pouco ainda se explora do desenvolvimento perceptual de bilíngues residentes em país de língua materna e, portanto, não

imersos na L2. Considerando a nova configuração de contextos de prática em língua estrangeira provenientes das tecnologias de informação e comunicação (TICs), é necessário investigar se o uso linguístico oral em L2 possibilitado por novos artefatos tecnológicos é capaz de alterar a representação e o processamento linguístico dessa natureza de forma significativa para aprendizes que residem em países de língua materna. Por exemplo, a grande quantidade de conteúdo e recursos interativos online (filmes, notícias, redes sociais, jogos etc) disponíveis em língua inglesa permitem a exposição à modalidade oral da língua, e até mesmo a interação conversacional com outros falantes da língua, sem a necessidade de presença física em países onde a língua inglesa prevalece socialmente.

Entendemos que a presente investigação vem ao encontro da necessidade de investigações sobre como ser usuário de uma L2 influencia a atividade e a arquitetura cognitiva, e como as línguas faladas são representadas na mente humana, especificamente no contexto do bilinguismo tardio e prevalentemente adquirido em situação de imersão em ambientes onde a L2 não é a língua socialmente dominante, como ocorre comumente entre aprendizes de L2 no Brasil. Estudos sobre a influência do bilinguismo na arquitetura mental e na organização neurocognitiva humana protagonizam as investigações em psicolinguística do bilinguismo na atualidade. Este protagonismo responde ao crescente interesse na comunidade científica por hipóteses acerca de vantagens cognitivas, neurais e sociais cuja observação em pessoas bilíngues vem sendo documentada, ocasionalmente com repercussões para além dos círculos estritamente acadêmicos (MARIAN, SHOOK, 2012).

Assim, o presente estudo visa trazer uma contribuição para o detalhamento da aquisição de L2 e do conseqüente processamento linguístico por bilíngues tardios brasileiros. Em nosso entendimento, este é um pré-requisito indispensável para garantir que qualquer debate sobre vantagens ou impactos cognitivos do uso de uma L2 por essa população não se assente em conclusões pautadas por observações empíricas de perfis de bilinguismo advindos de contextos sociolinguísticos significativamente díspares. Em nossa percepção tal estado de coisas fragilizaria essas conclusões e as tornaria no máximo especulativas.

A seguir abordaremos a caracterização do fenômeno linguístico que constitui o foco da presente investigação, que é um contraste fonotático das consoantes nasais /m/ e /n/ na organização silábica do inglês e do português que acarreta impactos sobre o valor fônico atribuído por leitores dessas línguas a grafemas compartilhados pelas convenções ortográficas presentes em ambas as línguas, e que ocorrem em posições grafotáticas análogas na ortografia de sílabas das duas línguas. Passamos, então, à exposição de um quadro teórico que propõe uma arquitetura de representação linguística baseada em uso de um paradigma experimental que revela algo da mecânica do reconhecimento

visual de palavras por leitores, discutindo sua conexão com o presente trabalho. Tratamos, em seguida, dos métodos e observações empíricas de nosso estudo, o que nos conduzirá às conclusões que ora apresentamos.

2. As consoantes nasais /m/ e /n/

2.1 As consoantes nasais /m/ e /n/ em língua inglesa

Em inglês, as consoantes nasais /m/ e /n/ ocorrem de forma distintiva (AVERY; EHRLICH, 1992; EWEN; HULST, 2001) em início e final de sílaba. Por exemplo, *meat* /mi:t/ ‘carne’ e *neat* /ni:t/ ‘arrumado’; *them* /ðem/ ‘eles, elas’ e *then* /ðen/ ‘então’, em que os pares de palavras se distinguem apenas pela consoante nasal na mesma posição silábica.

Embora o inglês seja considerado uma língua opaca em geral, há grande transparência ortográfica com relação aos sons que nortearam a nossa seleção de itens lexicais: consoantes nasais /m/ e /n/ em final de sílaba. Além disso, tentamos manter a transparência também no início da sílaba dos pares de palavras alvo (*them* e *then*). Em alguns pares, não foi possível a transparência total no início da sílaba, tal como em: *team* /tim/ ‘time’ e *teen* /tin/ ‘adolescente’. Para efeitos estritamente descritivos, os correlatos ortográficos presentes nos nossos itens lexicais alvo estão exemplificados na Tabela 1:

Tabela 1 - Relação entre as consoantes nasais /m/ e /n/ em final de palavra e seus correlatos ortográficos

Sons	Correlatos Ortográficos
/m/	<m>: <i>them</i> ‘eles, elas’ <me>: <i>some</i> ‘algum’ <mb>: <i>comb</i> ‘pente’
/n/	<n>: <i>then</i> ‘então’ <ne>: <i>line</i> ‘fila’

Fonte: elaboração própria

2.2 As consoantes nasais /m/ e /n/ no Português Brasileiro

Em PB, há distinção fonêmica das consoantes nasais /m/ e /n/ em início de sílaba assim como no inglês. Porém, a distinção fonêmica entre /m/ e /n/ no final de sílaba está neutralizada. Segundo Gontow *et al.* (2006, p. 61), a consoante nasal /m/ não ocorre em posição final no português brasileiro. Palavras como ‘sem’ são escritas com um ‘m’ em posição final, mas são pronunciadas como um ditongo nasal final [ẽi]. A consoante nasal /n/ também não ocorre na posição final do português brasileiro. Por exemplo, na primeira sílaba da palavra ‘inseto’ [ĩ]seto, pronunciamos uma vogal nasal e não uma vogal seguida de consoante nasal.

De acordo com Silveira (2012, p.15), as consoantes nasais grafadas em final de sílaba não têm característica articulatória consonantal em PB, pois ocorre a ditongação da vogal precedente e assimilação do traço nasal (ex.: bom /boN/). No PB, ocorre apenas o contraste entre vogal oral e nasal como em ‘vi’ /vi/ e ‘vim’ /viN/.

A Tabela 2 apresenta a relação entre o arquifonema /N/ em final de sílaba e seus possíveis correlatos ortográficos em final de palavra. Na visão estruturalista, o arquifonema é uma unidade fonológica que engloba diferentes fonemas de uma língua que são distintivos em certas posições, mas neutralizam essa distinção em outras posições (MATTOSO CÂMARA, 1998 [1970]). Por exemplo, as consoantes nasais /m/ e /n/ são distintivas em início de sílaba (ex.: /m/ão e /n/ão), mas estão neutralizadas em final de sílaba e representam apenas a nasalidade da vogal antecedente. É importante ressaltar que o grafema ‘n’ não ocorre em final de monossílabos e que sua ocorrência é também restrita no final de palavras em geral.

Tabela 2 - Relação entre a nasal em final de palavra e seus correlatos ortográficos

Sons	Correlatos Ortográficos
/N/	<m>: <u>sim</u> <n>: hífe <u>n</u> <~>: lã

Fonte: Adaptação baseada em Barroso (1996, p. 276)

3. Representações Linguísticas Derivadas do Uso e a Mecânica do Reconhecimento Visual de Palavras Faladas

Nesta seção, apresentamos a Teoria de Exemplos que motivou a nossa hipótese principal e em seguida, a mecânica do reconhecimento visual de palavras revelada pelo Paradigma Mundo-Visual, sendo este o paradigma experimental que deu suporte para a nossa tarefa de reconhecimento de palavra falada.

3.1 A Teoria de Exemplos

As principais abordagens referentes à aprendizagem de segunda língua envolvem o princípio de que a exposição a estímulos da língua alvo seja necessária, isto é, podem ser rotuladas como baseadas no uso. Uma delas é a Teoria de Exemplos (JOHNSON, 1997, 2005; PIERREHUMBERT, 2001, 2003) em que as representações mentais de L1 se reorganizam para a acomodação dos exemplos de L2. Desse modo, a frequência de uso linguístico de L2, ou a relação entre quantidade e qualidade de estímulo, é responsável pela aprendizagem e representação linguística de L2.

Na tentativa de explorar uma representação do conhecimento fonológico que seja um pouco mais próxima da riqueza da linguagem tal como experienciada e armazenada por falantes nativos ou não nativos, a questão central da Teoria de Exemplos é a hipótese de que as instâncias de uso impactem gradualmente na representação cognitiva da linguagem e que a estrutura linguística emerge do próprio uso (BYBEE, 2001, 2010). Desde os primeiros estudos psicolinguísticos, surgem questões sobre a organização do léxico mental e sobre o acesso lexical em sujeitos bilíngues, uma vez que falantes de duas ou mais línguas precisam não apenas armazenar itens de línguas diferentes, como também os distinguir tanto em processos de compreensão como de produção da linguagem.

A Teoria de Exemplos sugere que cada realização de palavra seja armazenada em uma espécie de nuvem de exemplos dentro do léxico mental. Por exemplo, quanto mais vezes e mais recentemente uma palavra for utilizada, mais forte será sua representação e mais rápido será o acesso ao exemplar da palavra. Para incorporar um novo exemplar, a nuvem se reorganiza para acomodar os novos padrões que foram atestados. Nesse sentido, a própria experiência ou uso linguístico em uma segunda língua delinear e reforçaria as representações no léxico mental e, conseqüentemente, diminuiria o tempo de competição para o acesso lexical no âmbito lexêmico.

Nas nuvens de exemplos, as palavras semelhantes ficam próximas e as divergentes ficam distantes. A unidade lexical tem propriedades fonéticas idiossincráticas, porém ela também é formada por estruturas sonoras que são compartilhadas com outras palavras. Considerando que as representações linguísticas não são apenas categorizações de eventos perceptuais descontextualizados, a Teoria de Exemplos sugere que todas as informações relativas ao item lexical estejam armazenadas e conectadas em nuvens de exemplos, conforme ilustra a Figura 1.

Figura 1 - Nuvem de Exemplos com a inclusão de fatores linguísticos e extralinguísticos



Fonte: Bybee (2001, p. 52)

Entretanto, o volume de fala que a pessoa processa em uma vida é tão alto que memórias individuais para cada uso de cada palavra são questionáveis. Logo, a Teoria de Exemplos responde

a essa questão de duas formas. A primeira é assumindo que a memória decaia e por isso, memórias de ontem sejam mais vívidas do que memórias de uma década atrás. A segunda é considerando que o exemplar individual (memória perceptual detalhada) não corresponda a uma experiência perceptual única, mas a uma classe equivalente de experiências perceptivas granulares.

Pierrehumbert (2001, p.3) propõe a extensão da Teoria de Exemplos de percepção para a produção, considerando que o uso linguístico real na comunicação envolve dados de percepção e produção que reforçam os exemplos. A produção é considerada tão sensível à força da representação quanto à classificação perceptual. Portanto, o modelo de produção seria uma extensão mínima dos trabalhos prévios de como os exemplos funcionam na percepção com a inversão da direção. O *feedback* ‘retorno’ da gramática fonológica no ciclo percepção-produção aprimora a capacidade de detectar limite de palavras e aumenta a atenção para detalhes fonéticos lexicalmente contrastivos (fonêmicos).

Em vista disso, com o objetivo de controlar e explicitar os parâmetros de uso e dominância linguística para além de delimitar proficiência e imersão, a justificativa que liga a Teoria de Exemplos ao acesso lexêmico por bilíngues ou aprendizagem de segunda língua é que os perfis de uso linguístico podem emergir da frequência de experiência em diferentes contextos comunicativos de L2 pelo falante imerso em país de L1. Os novos contextos de prática em língua estrangeira, tais como os provenientes das tecnologias de informação e comunicação e os novos artefatos tecnológicos, parecem promover uma nova configuração de dominância linguística na realidade brasileira (VALADARES, 2017). Nesse sentido, o aumento da exposição à língua adicional e a diversificação dos cenários onde ela acontece podem afetar o processamento linguístico de L2 por brasileiros porque o modo como as línguas são representadas e processadas depende não somente das propriedades formais do sistema linguístico como também das características cognitivas e interações dos indivíduos bilíngues com o meio em que vivem.

3.2 O Paradigma Mundo-Visual

Há mais de vinte anos, pesquisadores têm usado o paradigma mundo-visual para melhor caracterizar o processamento da fala com relação à competição semântica, visual e fonológica (TANENHAUS *et al.*, 1995; TANENHAUS; SPIVEY-KNOWLTON, 1996). O paradigma mundo-visual é uma técnica de rastreamento ocular que explora o fato de as pessoas fazerem movimentos sacádicos para objetos reais ou figuras de objetos em tela quando os nomes dos objetos

4 O movimento sacádico é de ambos os olhos entre dois ou mais pontos de fixação na mesma direção. Ele é rápido e simultâneo.

são mencionados em instruções orais. Os movimentos oculares refletem a mudança de atenção entre referentes, sendo rápidos, incrementais e antecipatórios. As fixações oculares, especificamente, estão probabilisticamente relacionadas à atenção. Consequentemente, a média de duração de fixação ocular geralmente é influenciada pela facilidade ou dificuldade de processamento. A duração da fixação é diretamente proporcional à dificuldade de processamento.

No paradigma mundo-visual, uma palavra alvo é apresentada em áudio e quatro imagens ou palavras escritas (um alvo, um ou mais competidores e um ou mais distratores) são apresentadas em tela enquanto ocorre o mapeamento dos movimentos oculares a partir de um rastreador ocular. Há, no entanto, três variantes no paradigma. Na primeira, o participante precisa olhar para a tela. Na segunda, o participante precisa clicar na palavra alvo. Na terceira, o participante precisa mover a imagem para algum outro lugar da tela. Apesar dessas diferenças, os resultados são semelhantes para as três variantes (HUETTIG; ALTMANN, 2004; YEE; SEDIVY, 2006; DAHAN; TANENHAUS, 2005, respectivamente).

A Figura 2 exemplifica o paradigma mundo-visual com a variante que utilizamos, aquela em que o participante precisa clicar na palavra ouvida. A palavra alvo que é apresentada em áudio é *them* /ðem/ ‘eles, elas’. No estímulo visual consta a palavra alvo: *them* /ðem/ ‘eles, elas’; a palavra competidora: *then* /ðen/ ‘então’; e as palavras distratoras: *doll* /da:l/ ‘boneca’ e *part* /pa:rt/ ‘parte’.

Figura 2 - Exemplo de tela de estímulo do experimento de reconhecimento de palavra baseado no Paradigma Mundo-Visual



Fonte: elaboração própria

Geralmente, as imagens ou palavras em tela recebem quantidade e intensidade de fixações equivalentes até que a palavra alvo seja pronunciada para que a atenção para o competidor semântico, visual ou fonológico supere os demais distratores e a palavra alvo seja selecionada. Caso os movimentos oculares não representassem o processamento natural da linguagem, as fixações seriam aleatórias (DAHAN; TANENHAUS, 2005).

A grande dificuldade dos estudos envolvendo o paradigma mundo-visual consistia em selecionar imagens alvo. No entanto, esse problema foi resolvido quando Huettig e McQueen (2007), McQueen e Viebahn (2007) e Weber *et al.* (2007) mostraram que podemos usar palavra escrita no paradigma mundo-visual, obtendo resultados equivalentes aos estudos com imagens, principalmente com relação às sobreposições fonológicas.

Segundo Barr (2008, p. 458), o diferencial do paradigma mundo-visual em relação às técnicas psicolinguísticas que envolvem tempo de reação (decisão lexical, sombreamento) é o potencial para acessar o processamento enquanto ele está ocorrendo. Ao envolver medidas de rastreamento ocular, esse paradigma oferece uma medida mais refinada do processamento lexical na fala contínua, e permite o uso de tarefas mais naturais do que medidas de resposta que requerem um julgamento metalinguístico. Considerando que o julgamento metalinguístico possa induzir a diferentes estratégias de processamento, uma medida anterior ao julgamento metalinguístico oferece dados mais diretos e com maior sensibilidade temporal.

A sensibilidade temporal do paradigma mundo-visual faz com que ele seja uma ótima opção dentre os métodos *online* ‘em tempo real’ da psicolinguística experimental (SEKERINA, 2014) porque as medidas de movimento ocular durante o reconhecimento de palavra falada são relacionadas ao tempo de forma precisa (MIRMAN; MAGNUSON, 2009). Segundo Liversedge e Findlay (2000, p. 9), a duração e a quantidade de fixações aumentam enquanto sacadas diminuem em função da semelhança entre palavras alvo e palavras competidoras. Nesse sentido, em uma análise de fixações oculares, temos que decidir quais fixações são somadas para detectar uma dificuldade de processamento. A soma pode favorecer a continuidade temporal (fixações ocorrendo no mesmo intervalo de tempo) e/ou espacial (fixações ocorrendo no mesmo local).

Os críticos do paradigma mundo-visual questionam os efeitos de frequência das palavras e dos competidores que não estão em tela para afirmar que não é o processamento natural da linguagem que está sendo observado pelo mapeamento ocular. Mas Dahan *et al.* (2001: 359) argumentam que não há restrição de competidores aos que estão em tela com base nas evidências de efeitos de frequência.

3.2.1 Paradigma Mundo-Visual: Associação entre Fonologia e Ortografia

A interação entre o estímulo ortográfico e acústico é complexa (BASSETI, 2008). Nesta seção, apresentamos, em primeira instância: Seidenberg e Tanenhaus (1979), Burton *et al.* (1993) e Ziegler e Ferrand (1998), que investigaram os efeitos da consistência ou inconsistência da correspondência grafema e fonema no tempo de processamento das palavras em tarefas psicolinguísticas. Esperava-se

que a forma ortográfica e a forma fonológica com correspondência direta de um para um fossem identificadas mais rapidamente do que uma forma ortográfica com múltiplas correspondências fonológicas ou uma forma fonológica com múltiplas correspondências ortográficas. Ou seja, esperava-se que quanto mais regular, mais fácil fosse a decodificação.

Seidenberg e Tanenhaus (1979) relataram que as respostas à pergunta ‘O estímulo rima?’ foram facilitadas apenas nas condições em que fonologia e ortografia eram compartilhadas (ex.: *pie* /paɪ/ ‘torta’ e *tie* /taɪ/ ‘gravata’). Seidenberg e Tanenhaus também encontraram que rimas ortograficamente semelhantes eram mais rapidamente reconhecidas do que as que eram diferentes tais como: *rye* /raɪ/ ‘centeio’ e *tie* /taɪ/ ‘gravata’. A condição de semelhança apenas ortográfica não foi testada por eles.

Burton *et al.* (1993) investigaram respostas de participantes em tarefas de sombreamento e decisão lexical a partir de alvos auditivos precedidos por *priming*⁵ oral fonologicamente relacionado (ex.: *clue* /klu/ ‘pista’ e *zoo* /zu/ ‘zoológico’), ortograficamente (ex.: *head* /hed/ ‘cabeça’ e *bead* /bid/ ‘conta’), ou ambos (ex.: *porch* /pɔ:htʃ/ ‘varanda’ e *torch* /tɔ:htʃ/ ‘tocha’) versus *priming* não relacionado com o alvo (ex.: *soap* /soʊp/ ‘sabão’ e *zoo* /zu:/ ‘zoológico’). Burton *et al.* encontraram que as respostas aos pares que compartilhavam informações fonológicas e ortográficas ou apenas informações fonológicas foram facilitadas em comparação com as condições não relacionadas. Os pares com relação somente ortográfica não se diferenciaram dos não relacionados.

No estudo de Ziegler e Ferrand (1998), itens consistentes (uma única forma de escrita) obtinham resposta mais rápida do que itens inconsistentes (formas de escrita variadas) em tarefa de decisão lexical. Nesse sentido, Seidenberg e Tanenhaus (1979), Burton *et al.* (1993) e Ziegler e Ferrand (1998) verificaram que existe um efeito facilitador da consistência entre grafema e fonema, implicando em menor tempo de reação nas tarefas. E também verificaram que existe um efeito facilitador de menor grau quando há somente relação fonológica. No entanto, a influência da relação estritamente ortográfica na competição não foi comprovada.

Desse modo, é relevante apresentar um estudo recente (SHEN *et al.*, 2018) que investigou se o efeito fonológico sobre a competição lexical poderia estar completamente dissociado do efeito ortográfico ao lidar com a língua chinesa, uma língua logográfica. A língua logográfica mapeia as palavras diretamente em unidades de significado ao invés de unidades fonêmicas como as línguas alfabéticas. Como a ligação entre ortografia e fonologia é relativamente fraca em chinês, a investigação da língua chinesa parece mais adequada para a possível distinção entre o papel da fonologia e o papel

⁵ Exposição prévia a um estímulo que pode exercer influência sobre um subsequente, sem que o indivíduo tenha consciência explícita sobre ele.

da ortografia no reconhecimento de palavra falada.

Shen *et al.* (2018) manipularam os itens alvo e competidores para conterem sobreposição fonológica total, parcial e nenhuma. A sobreposição fonológica ocorria em diferentes posições na sílaba. Shen *et al.* (2018) concluíram que os competidores fonológicos atraíram mais fixações e a competição foi proporcional ao grau de sobreposição fonológica, conforme era esperado. Os resultados vindos de uma língua logográfica e não alfabética são muito relevantes para evidenciar que o efeito de competição fonológica pode ocorrer sem a influência da ortografia.

Nesse sentido, existem evidências robustas de que ocorra o efeito de competição fonológica no acesso lexical com a apresentação de estímulo auditivo e visual. Contudo, o efeito ortográfico parece ocorrer apenas a partir da correspondência com o fonológico. Sendo assim, quanto maior a transparência ortográfica (correspondência grafema e fonema) da língua, menor será o tempo de reação em tarefas psicolinguísticas que investiguem o acesso lexical.

Há muito ainda para se descobrir sobre a relação entre ortografia e fonologia, considerando os diferentes tipos de escrita e que algumas línguas alfabéticas são mais transparentes do que outras. Como vimos anteriormente, o grau de transparência entre grafema-fonema das línguas pode influenciar no tempo para reconhecimento de palavra.

4. Metodologia

4.1 Material

No total, foram 60 telas de estímulos com quatro palavras para cada participante. Dentre elas, 15 continham os pares experimentais (alvo e competidor) e dois distratores. 45 telas eram compostas apenas por distratores. Foram construídas duas versões do experimento de modo que as palavras dos pares experimentais fossem apresentadas em alternância para cada participante. Cada participante era exposto a uma palavra do par experimental apenas uma vez. Portanto, uma palavra alvo de um dos pares para um dado participante era a palavra competidora na versão do experimento administrada para o participante seguinte

Ainda sobre os pares experimentais, quinze palavras monossílabas terminadas na consoante nasal /m/ e quinze terminadas na consoante nasal /n/ foram selecionadas para formarem pares que compartilhassem o início da sílaba, tentando manter a maior transparência ortográfica possível. Por exemplo, o par mínimo: *them* /ðem/ ‘eles, elas’ e *then* /ðen/ ‘então’ que só se distingue pela consoante nasal final. Cada tela de estímulo de itens experimentais recebia também um par de distratores sem

relação fonológica ou semântica com os alvos (ex.: *doll* /dɒl/ ‘boneca’ e *part* /pɑːt/ ‘parte’), porém o tamanho (quantidade de sílabas e número de letras por tela de estímulo) era controlado. Os distratores foram extraídos de um *website* (*Random Lists*) que gera palavras aleatórias.

Além dos 15 itens experimentais compostos por dois itens em competição e dois distratores cada, acrescentamos três grupos de 15 itens distratores completos. Os três grupos também continham um par de palavras que incitavam competição (por homofonia, rima e relação de sinonímia aproximada, antonímia e/ou pertencimento ao mesmo campo semântico) para evitar que os participantes tomassem conhecimento dos nossos objetos de estudo específicos. Cada par em competição distratora era completado com dois distratores sem relação fonológica ou semântica. Um exemplo de item distrator completo é composto pelo quarteto: *flour* /flaʊər/ ‘farinha’, *flower* /flaʊər/ ‘flor’, *mouth* /maʊθ/ ‘boca’, e *hammer* /ˈhæm.ər/ ‘martelo’.

4.2 Participantes

4.2.1 Falantes Nativos de Inglês

Os participantes do grupo controle foram nove falantes nativos de inglês de diferentes nacionalidades (7, Estados Unidos; 1, Nova Zelândia e 1, Gana). Dentre eles, seis homens e três mulheres. As idades eram de 18 a 72 anos (média de 41 anos e desvio padrão de 20,46). Eles tinham visão normal ou corrigida e não tinham conhecimento sobre o que estava sendo investigado. A grande diferença de idade entre os participantes favorece a diversidade da nossa amostra para que a variável preditora de comportamento linguístico seja apenas o fato de serem falantes nativos de língua inglesa.

4.2.2 Falantes Não-Nativos de Inglês

Os participantes foram, inicialmente, 31 estudantes de graduação de 16 a 42 anos (média de 22,9 anos e desvio padrão de 5,75). Dentre eles, oito homens e 23 mulheres. Todos eram falantes nativos de PB e aprendizes de inglês como L2. Eles tinham visão normal ou corrigida e não tinham conhecimento sobre o que estava sendo investigado. Eles foram pré-selecionados pelo *Vocabulary Levels Test* ‘Teste de Amplitude Lexical’ (*VLT - NATION, BEGLAR, 2007*) em inglês como L2. Três participantes foram eliminados por obterem pontuação abaixo do que é considerado nível 1 (conhecimento das 2000 palavras mais frequentes da língua) e portanto, poderiam fazer as tarefas tentando adivinhar as respostas ao invés de realmente evidenciarem os esforços cognitivos durante o processamento de L2. Portanto, todas as análises foram feitas considerando 28 participantes não nativos de língua inglesa.

4.3 Procedimentos

A coleta de dados⁶ consistiu em quatro etapas. Os resultados das quatro etapas foram combinados para a verificação das hipóteses experimentais. A primeira etapa foi um Teste de Amplitude Lexical em L2 (VLT – NATION, BEGLAR, 2007) temporalizado em 10 minutos para seleção de participantes bilíngues do par linguístico PB-Inglês e posterior análise de dados com relação aos diferentes níveis de amplitude lexical. O teste mede o tamanho do vocabulário do aprendiz a partir do conhecimento do significado de palavras de diferentes grupos de frequência e a pontuação varia entre zero e 90 pontos.

A segunda etapa foi um Questionário de Levantamento de Usos Linguísticos (VALADARES, 2017) do qual foram extraídos apenas os itens de uso linguístico oral em L2 para a confecção do perfil de cada participante. Os itens relacionados à seleção de língua (Português ou Inglês) em que atividades envolvendo oralidade são realizadas foram: a) pesquisa por informações na internet cujo conteúdo é falado; b) procuro assistir a filmes, seriados e outros programas; c) escuto músicas; d) jogo *games* nos quais há falas; e) tento aprender a letra de músicas das quais gostei; e f) converso com pessoas que falam a língua fluentemente.

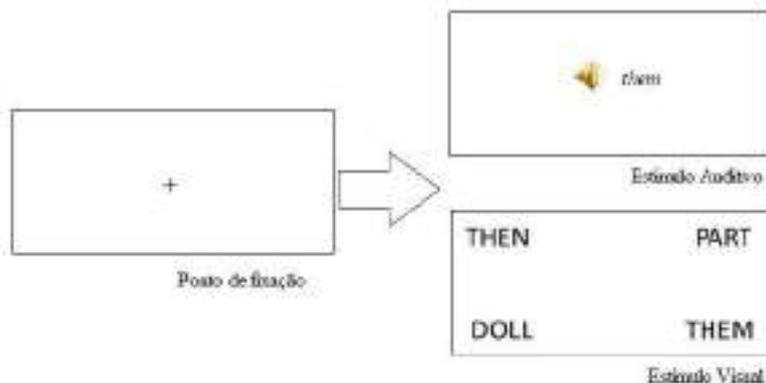
O participante respondia em uma escala de 0 a 4: 0 (não faço essa atividade ou somente em português); 1 (mais em português do que inglês); 2 (igualmente em português e inglês); 3 (mais em inglês do que português) e 4 (somente em inglês). Ao final do questionário, os participantes obtinham um valor de 0 a 24 que os agrupava em diferentes grupos de uso linguístico oral em L2. Quanto maior a pontuação, maior o uso linguístico oral de L2.

A terceira etapa foi o experimento de reconhecimento de palavra falada baseado no Paradigma Mundo-Visual. Os áudios foram apresentados aos participantes por caixas de som ligadas ao computador de apresentação dos estímulos e os movimentos oculares foram registrados pelo rastreador ocular *Eyelink 1000*. Cada tela de estímulo continha quatro palavras escritas (uma em cada canto da tela). A posição das palavras alvo e competidoras foi balanceada ao longo das tarefas e a ordem das telas de estímulos foi aleatorizada. Os estímulos visual e auditivo eram apresentados simultaneamente, pois quando o estímulo visual é escrito não é necessária distância temporal entre áudio e tela para que haja competição fonológica (HUETTIG; MCQUEEN, 2007). Além disso, a seleção de um item era esperada apenas após o *offset* ‘final’ da palavra. O estímulo visual fica em tela até o participante selecionar uma das palavras através de um clique com o *mouse*.

6 Os dados de todas as etapas estão disponíveis em: http://bit.ly/dados_tese_Rodrigues_2019 e a tabela com os dados tabulados está no Apêndice.

Entre cada tela de estímulo, aparecia uma cruz preta no centro da tela para que o participante fixasse o olhar e tivesse um ponto de partida fixo para o movimento ocular e para que correções necessárias com relação ao mapeamento do olhar fossem efetuadas. O esquema de apresentação de estímulos está ilustrado na Figura 3, que apresenta o ponto de fixação que aparecia antes de cada tela de estímulo composta por quatro palavras escritas e um estímulo auditivo (palavra isolada).

Figura 3 - Sequência de Apresentação de Estímulos do Experimento de Reconhecimento de Palavra Falada



Fonte: elaboração própria

Ao comparecerem ao laboratório, os participantes eram questionados com relação à visão e à audição, e apresentados ao rastreador ocular. Eram então solicitados a fixar a cabeça em um descanso de queixo e testa para minimizarem o movimento durante o experimento. Após ajustar a altura da cadeira com relação ao descanso de queixo e testa, a calibragem do rastreador era efetuada. Durante a calibragem, o participante fixava o olhar em nove pontos em tela que mudavam de lugar e a margem de erro aceitável entre a distância da projeção da pupila e do ponto de fixação apresentado em tela era de no máximo 1 grau. Os participantes foram instruídos a clicar com o *mouse* na palavra alvo assim que a identificassem. Houve uma etapa de treinamento com seis telas de estímulos distratores para a familiarização com o tipo de tarefa. Em seguida, foram apresentados em ordem aleatória 15 itens experimentais e 45 distratores.

A quarta tarefa foi de leitura de palavra em voz alta. A palavra isolada era disposta no centro da tela e permanecia por um segundo. Além das 30 palavras alvo, mantivemos 90 distratores. As 120 palavras foram apresentadas de forma aleatória. O propósito dessa tarefa foi verificar a produção das consoantes nasais /m/ e /n/ em final de palavra de maneira contrastiva para discutir a interação entre produção e percepção. As análises buscavam traços consonantais no final da sílaba e/ou diferença no valor de F2, assim como foi feito por Becker (2014, p. 112). Desse modo, as análises de oitiva não coincidentes entre duas pesquisadoras eram verificadas pela análise acústica comparativa dos pares de palavras em competição do mesmo falante e com a produção dos falantes nativos. Os participantes

recebiam uma nota 0 ou 1. Quando a produção da consoante nasal /m/ não era distinta da produção da consoante nasal /n/ em final de palavra ou era muito divergente da produção dos falantes nativos de inglês, o participante recebia pontuação 0. Quando a produção da consoante nasal /m/ era distinta da produção da consoante nasal /n/ em final de sílaba ou era convergente com a produção dos falantes nativos de inglês, o participante recebia pontuação 1.

4.4 Contextualização e Critérios para as Medidas de Rastreamento Ocular utilizadas para as Análises Descritivas, Gráficas e Estatísticas

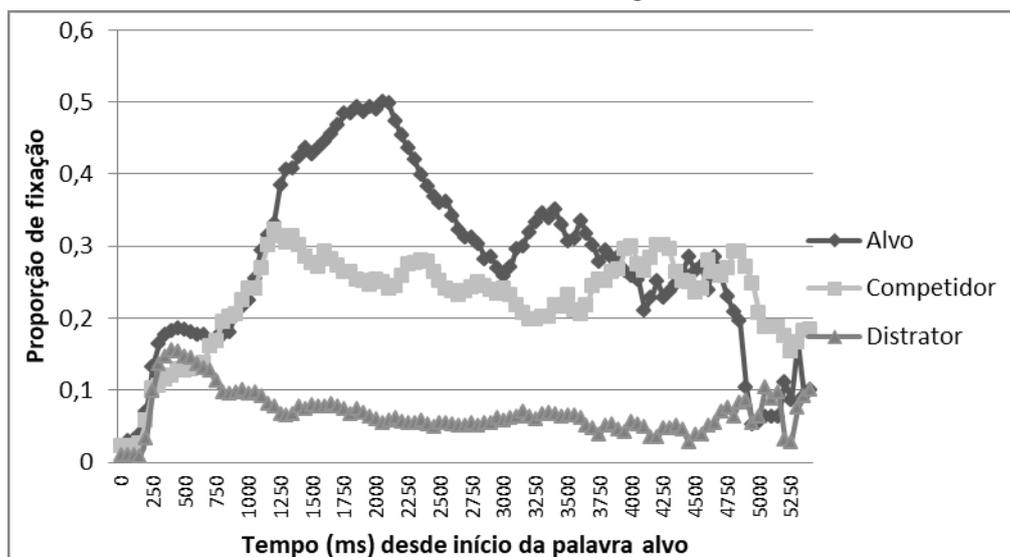
Segundo Tanenhaus e Trueswell (2006, p. 875), a maioria dos estudos que utilizam os movimentos oculares para investigar a leitura faz uso dos movimentos oculares como medida de carga de processamento. A principal medida dependente é a duração de fixação ocular. A hipótese de ligação entre duração de fixação e processos subjacentes é que o tempo de leitura aumenta quando o processamento se torna mais difícil. Em contraste, a maioria dos estudos do paradigma mundo-visual utiliza os movimentos oculares como uma medida representacional. A medida dependente primordial é quando e onde o sujeito fixou o olhar enquanto a palavra estava sendo ouvida.

Na nossa tarefa de reconhecimento de palavra falada, como as áreas de interesse tinham o mesmo tamanho e circundavam as palavras escritas, a identificação do clique na palavra correta e as fixações oculares foram medidas por cada área de interesse. Os erros de identificação do item lexical não foram excluídos da análise porque demonstravam ainda mais a competição fonológica que ocorre entre palavras alvo e competidoras para o par linguístico aqui estudado.

Nesse sentido, Tanenhaus e Trueswell (2006, p. 874) afirmaram ser útil fazer essa distinção entre medidas comportamentais de processamento linguístico que medem a dificuldade de processamento e medidas que exploram representações. Porém, a distinção é mais heurística do que categórica porque muitas medidas de resposta combinam aspectos de ambas e a natureza do objeto de estudo é que determina qual o tipo de medida de resposta apropriado.

A Análise de Crescimento da Curva (MIRMAN *et al.*, 2008; SALVERDA; TANENHAUS, 2010) é uma metodologia de análise visual amplamente utilizada para experimentos baseados no paradigma mundo-visual (TANENHAUS *et al.*, 1995; TANENHAUS, SPIVEY-KNOWLTON, 1996). A curva mostra o percurso temporal da proporção de fixação para palavras alvo, competidoras e distratoras. Utilizamos janelas temporais em intervalos de 100ms tal como sugere a literatura. Um exemplo da Análise de Crescimento da Curva pode ser visto no Gráfico 1.

Gráfico 1 - Proporção de fixação para todos os estímulos até a seleção de uma das palavras em tela para falantes não nativos de inglês



Fonte: elaboração própria

O eixo das abscissas do Gráfico 1 corresponde ao tempo total em milissegundos até a seleção da palavra alvo. O eixo das ordenadas corresponde à porcentagem ou proporção do tempo de fixação ocular por área de interesse (alvo, competidor e média dos dois distratores) com relação ao tempo total de fixação por tela de estímulo. A soma da porcentagem ou proporção de duração das fixações oculares para as quatro áreas de interesse não é igual a 100% porque as fixações oculares para outras regiões da tela durante a apresentação do estímulo ficam fora do cálculo. No caso dos falantes não nativos de inglês, foram 420 telas de estímulos alvo (15 telas para 28 participantes).

Tanenhau e Trueswell (2006, p. 869) questionam o fato da curva de crescimento aparentemente mostrar que a proporção das fixações oculares seja uma medida contínua. Os autores ressaltam que é mais apropriado entender os movimentos oculares como uma aproximação a uma medida contínua, porque a pessoa não vai simplesmente aumentando a fixação para o mesmo local, ela faz movimentos sacádicos para outros lugares e também movimentos de regressão que ‘interrompem’ as fixações oculares em lugares específicos.

Para a nossa análise estatística, de acordo com a natureza da questão sobre investigação – a intensidade de competição – a medida apropriada precisou da junção das medidas comumente utilizadas para leitura e para o paradigma mundo-visual. Ou seja, a média do tempo total de fixação em alvo e competidor ao longo de intervalos temporais e, em alguns casos, a diferença entre essas duas medidas. A análise por diferentes janelas temporais, inicialmente, não é relevante porque o nosso estímulo auditivo consiste apenas na palavra alvo isolada, porém, ao separarmos o tempo de fixação

em áreas de interesse por janelas temporais idênticas e sequenciais aumentamos a continuidade da medida. Dessa forma, a medida relaciona a representação e o processamento porque medidas de carga de processamento acessam mudanças temporárias na complexidade do processo, e depois usam essas mudanças para fazer inferências sobre os processos e representações subjacentes (TANENHAUS, TRUESWELL, 2006).

A nossa hipótese principal se baseia no impacto do perfil de uso linguístico oral de inglês-L2 no âmbito lexêmico do acesso lexical. Os perfis de uso foram relacionados com o tempo de fixação ocular como medida de intensidade de processamento. O maior tempo de fixação para alvo e competidor, assim como a menor diferença entre tempo de fixação para alvo e competidor indicam um custo maior de processamento ou intensidade de competição. E vice-versa, quanto menor o tempo de fixação para alvo e competidor e maior a diferença entre o tempo de fixação para alvo e competidor, menor o custo de processamento ou intensidade de competição.

5. Análises

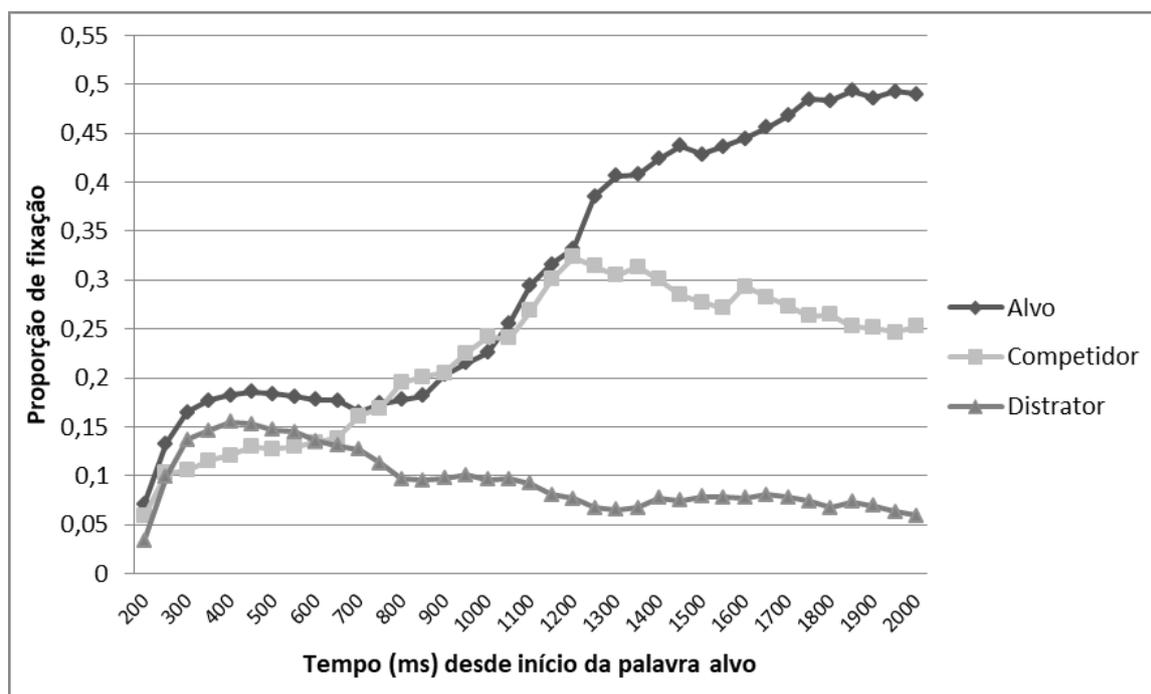
De acordo com o primeiro desdobramento da hipótese principal, esperava-se encontrar uma diferença significativa entre fixações oculares para palavras alvo e competidoras em relação às distratoras. Também que a fixação ocular para alvo e competidores fosse maior para falantes não nativos do que para falantes nativos de língua inglesa.

Primeiramente, foi feita uma Análise de Crescimento da Curva (MIRMAN *et al.*, 2008; SALVERDA; TANENHAUS, 2010) que mostra a proporção de fixação em área de interesse (alvo, competidor e distratores) ao longo do tempo para todos os participantes e todas as palavras até a seleção de uma das quatro palavras em tela.

Segundo Tanenhaus e Trueswell (2006, p. 876), a informação sobre o percurso temporal das fixações está diretamente ligada à lógica do Paradigma Mundo-Visual porque permite grandes inferências sobre qual informação no estímulo auditivo possa ter atraído a fixação para o estímulo visual. Além do Gráfico 1, que mostra o crescimento da curva com o tempo total até a seleção da palavra alvo, apresentamos os Gráficos 2 e 3 com o limite temporal de 2000ms para facilitar a visualização do momento em que as fixações oculares para os distratores começam a diminuir e a competição entre os pares de alvo e competidor tem início. Além disso, é possível visualizar o momento em que a proporção das fixações oculares para palavras alvo se distancia da proporção das

fixações oculares para as competidoras.

Gráfico - Proporção de fixação até 2000ms para falantes não nativos de inglês

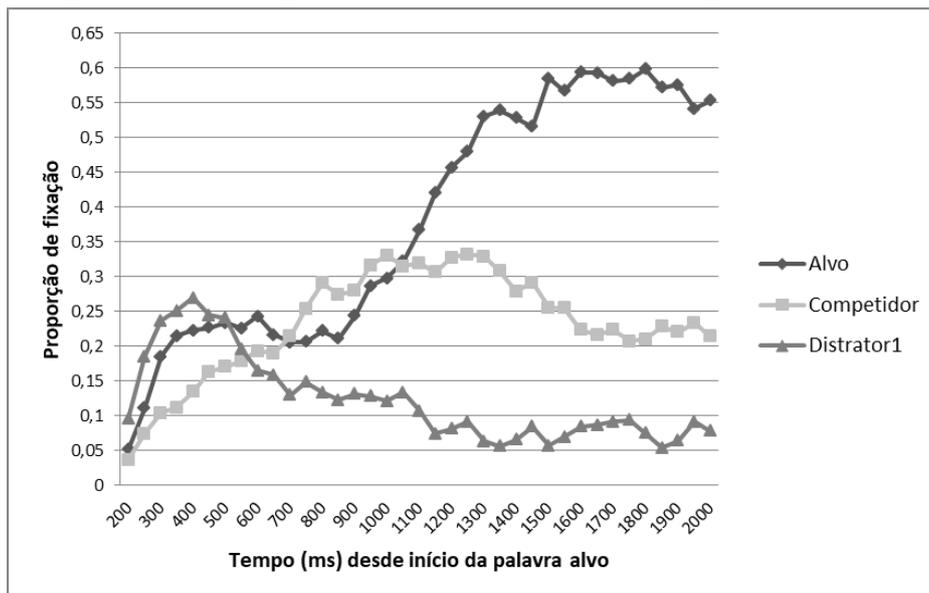


Fonte: elaboração própria

O eixo das abscissas do Gráfico 2 corresponde aos primeiros 2000 ms antes da seleção da palavra com o *mouse* pelo participante. O eixo das ordenadas corresponde à porcentagem ou proporção do tempo de fixação ocular por área de interesse (alvo, competidor e média dos dois distratores) com relação ao tempo total de fixação por tela de estímulo. A soma da porcentagem ou proporção de duração das fixações oculares para as quatro áreas de interesse não é igual a 100% porque as fixações oculares para outras regiões da tela durante a apresentação do estímulo ficam fora do cálculo. No caso dos falantes não nativos de inglês, foram 420 telas de estímulos alvo (15 telas para 28 participantes).

No Gráfico 3, apresentamos os dados dos nove falantes nativos de inglês em um total de 135 telas de estímulos alvo (15 telas para 9 participantes). Ao compararmos os Gráficos 2 e 3, percebemos que o distanciamento dos distratores ocorre por volta do 650ms em ambos. E notamos que a proporção de fixação para alvo e competidor por falantes não nativos de inglês a partir de 650ms apresentam valores mais próximos do que por falantes nativos de inglês. A diferenciação entre alvo e competidor para o grupo de não nativos ocorre por volta dos 1200 ms. Para o grupo de nativos, a diferenciação ocorre por volta dos 1000 ms, ou seja, 200 ms mais cedo.

Gráfico 3 - Proporção de fixação para todos os estímulos até 2000ms para falantes nativos de inglês



Fonte: elaboração própria

Em segundo lugar, seguimos um procedimento de cálculo de razão introduzido por Salverda *et al.* (2007, p. 5) para examinar os efeitos de competição e testar se a diferença de fixação ocular em competidor versus distratores foi significativa. O cálculo da razão é a divisão entre a proporção de fixação para o competidor e a soma da proporção de fixação para o competidor e a média da proporção de fixação para os distratores.

Segundo Salverda *et al.* (2007, p. 5), uma razão de 0,5 indicaria que o competidor não atraiu mais fixações do que distratores; uma razão menor do que 0,5 indicaria que os distratores atraíram mais fixações do que o competidor e uma razão maior do 0,5 indicaria que o competidor atraiu mais fixações do que os distratores.

Tabela 3: Proporções totais de fixação para áreas de interesse (alvo, competidor e distratores) para o grupo de falantes não nativos e de falantes nativos de inglês

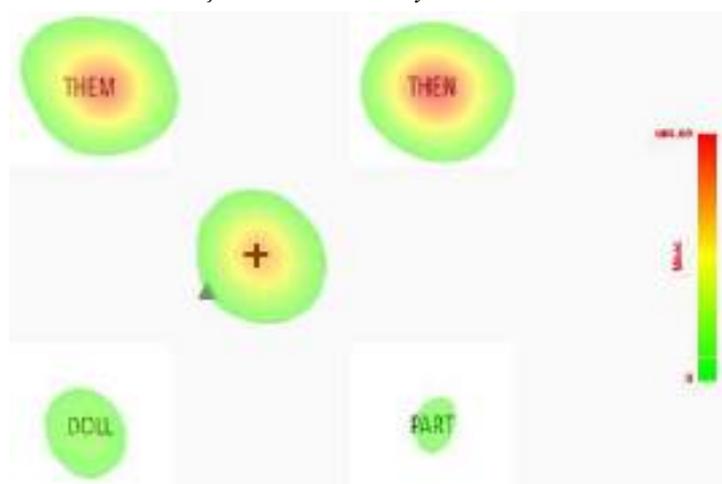
	Falantes Não Nativos de Inglês	Falantes Nativos de Inglês
Alvo	32,94	38,6
Competidor	24,03	21,06
Distrator	7,51	8,45
Distrator	7,4	8,73
Razão entre competidor e distratores	0,76	0,71

Fonte: elaboração própria

A Tabela 3 apresenta a proporção de fixação total para cada área de interesse por grupo (falantes não nativos e falantes nativos de inglês), e o resultado do cálculo da razão entre competidor e distratores. Considerando que a Tabela 3 mostra uma razão de 0,76 para os falantes não nativos de inglês e uma razão de 0,71 para os falantes nativos de inglês, a fixação em competidores é significativamente maior do que em distratores para ambos os grupos. Além desta diferença significativa, um segundo desdobramento da hipótese principal era que falantes não nativos de inglês teriam, de modo geral, maior dificuldade em distinguir as consoantes nasais /m/ e /n/ em final de sílaba do que os falantes nativos de inglês.

Adicionalmente, a Figura 4 apresenta um mapa de calor dos movimentos oculares de um falante não nativo de inglês ao olhar para uma tela de estímulos alvo enquanto escutava a palavra *them*. A figura mostra que houve maior quantidade (tamanho) e tempo (cor) de fixação para o par de palavras em competição lexêmica, o que indica um grande custo de processamento para decidir qual das duas era a palavra alvo.

Figura 4: Mapa de calor das fixações oculares para um falante não nativo de inglês extraído da sessão de visualização de dados no *EyeLink DataViewer*



Fonte: elaboração própria

Na sequência, foi feito o teste de normalidade *Shapiro-Wilk* dos dados de tempo de fixação ocular para palavras alvo e competidoras no grupo de nativos e não nativos de inglês. Se o valor de $p > 0,05$, a distribuição é normal e o teste correto para comparar os dois grupos é o teste t para amostras independentes. O teste t mostrou que estatisticamente existe diferença em ambos os casos ($p < 0,0001$), e que o grupo de não nativos tem um maior tempo de fixação ocular do que os nativos tanto no alvo quanto no competidor. A Tabela 4 apresenta os resultados do teste t:

Tabela 4 - Dados do teste t de comparação entre: a) o tempo de fixação em palavra alvo para o grupo de falantes nativos e para o grupo de falantes não nativos; b) o tempo de fixação em palavra competidora para o grupo de falantes nativos e para o grupo de falantes nativos de inglês

	Grupo	n	Média	Desvio padrão	t	gl	p valor
Alvos	Nativos	9	17,25	2,90	-15,465	35	< 0,0001
	Não nativos	28	51,56	6,40			
Competidores	Nativos	9	9,88	1,70	-19,456	35	<0,0001
	Não nativos	28	35,51	6,29			

Fonte: elaboração própria

A não rejeição do primeiro desdobramento da hipótese principal demonstra que o paradigma mundo-visual (TANENHAUS *et al.*, 1995; TANENHAUS; SPIVEY-KNOWLTON, 1996) possibilita a obtenção de dados de competição lexical durante o acesso lexical no âmbito lexêmico. Com relação ao segundo desdobramento da hipótese principal, a maior fixação para alvo e competidor para falantes não nativos de inglês demonstra que os falantes não nativos gastam mais tempo de processamento para selecionar entre um par de palavras em competição lexêmica. É interessante ressaltar que mesmo com a grande diferença de idade entre falantes não nativos e nativos de inglês, ainda foi possível visualizar a diferença estatística entre os dois grupos.

É sugerido que tal dificuldade possa estar relacionada com o fato de o brasileiro não fazer distinção fonêmica das consoantes nasais /m/ e /n/ em final de sílaba na sua língua materna e precisar da emergência desse som a partir do uso linguístico oral de L2 para o armazenamento e aumento de força dos exemplares de palavras com tais sons.

De acordo com a hipótese principal, esperava-se que o fator sobressalente de impacto sobre a intensidade de competição fosse o perfil de uso linguístico oral pela relação direta com a percepção no reconhecimento de palavra falada. Nesse sentido, esperava-se que a amplitude lexical e a precisão da produção oral afetariam a competição no âmbito lexêmico do acesso lexical com menor intensidade do que o perfil de uso linguístico oral em L2.

Foram realizados três diferentes modelos de regressão nos quais se desejava testar a influência das variáveis independentes (Amplitude Lexical em L2, Precisão na Produção Oral em L2, e Perfil de Uso Linguístico Oral em L2) sobre a diferença entre o tempo de fixação ocular em palavras alvo e palavras competidoras, ou seja, sobre a intensidade de competição lexêmica extraída do *EyeLink DataViewer*. O programa estatístico utilizado foi o SPSS, versão 24. O resultado dos modelos está exposto na Tabela 5:

Tabela 5 - Modelos de regressão univariados

Modelo	Variáveis Independentes	Coefficiente (B1)	Nível de explicação total (R2)	Significância do modelo (valor p)
1	Amplitude Lexical em L2	0,38	22,10%	0,007
2	Tarefa de Produção em L2	1,21	12,10%	0,039
3	Uso Linguístico Oral em L2	2,07	25,80%	0,003

Fonte: elaboração própria

O primeiro modelo verificou o impacto da amplitude lexical sobre a intensidade de competição. O modelo de regressão linear mostrou que o coeficiente foi 0,38, ou seja, a amplitude exerce uma influência positiva sobre a variável dependente, a significância do modelo foi $p=0,007$ e o nível de explicação total foi 22,1%. Isso significa que 22,1% da variação presente em duração de fixação ocular para alvo e competidor é explicada por amplitude lexical, que é, portanto, uma medida relacionável ao uso linguístico oral, apesar de ser uma medida de escrita.

O segundo modelo verificou o impacto da acurácia na produção oral em L2 sobre a intensidade de competição. Esperava-se que a maior precisão na produção das consoantes nasais /m/ e /n/ em final de sílaba em inglês-L2 explicasse a menor competição no reconhecimento de palavra falada. Ou seja, a acurácia na produção, por depender de uma representação fonológica da palavra razoavelmente estável e suficiente para diferenciá-la de um par mínimo, pode ser o suporte da maior rapidez no acesso lexical na operação de reconhecimento auditivo da palavra. O modelo de regressão linear mostrou que o coeficiente foi 1,21, a significância do modelo foi $p=0,039$ e o nível de explicação total foi 12,1%. Ou seja, por ter um coeficiente positivo é possível afirmar que quando se aumenta a acurácia na tarefa de produção, a diferença entre fixação para alvo e competidor aumenta e a intensidade de competição diminui. Tal resultado demonstra que a percepção e produção estão diretamente ligadas tal como sugerido pela Teoria de Exemplares.

O terceiro modelo verificou o impacto do uso linguístico oral em L2 sobre a intensidade de competição. O modelo de regressão mostrou que o coeficiente foi 2,07, a significância do modelo foi $p=0,003$ e o nível de explicação total foi 25,8%. O uso linguístico oral em L2, portanto, exerce influência positiva, uma vez que o coeficiente é positivo e significativo, sobre a organização do sistema sonoro do indivíduo tal como previsto pela Teoria de Exemplares (JOHNSON, 1997, 2005; PIERREHUMBERT, 2001, 2003) como um modelo baseado no uso (BYBEE, 2001, 2010).

Por conseguinte, constatou-se que o perfil de uso linguístico oral é o fator sobressalente de

impacto sobre a intensidade de competição, e que quanto maior a amplitude lexical em L2 e/ou mais precisa a produção oral em L2, menor é a competição no âmbito lexêmico do acesso lexical.

Considerações finais

A análise mostrou que o rastreamento ocular permite visualizar e quantificar a intensidade da competição no âmbito lexêmico do reconhecimento de palavras escritas após exposição de sinal acústico pela sensibilidade temporal dos dados de fixações oculares. O percurso de competição apresentado pelas curvas de crescimento mostra que as fixações para palavra alvo, competidora e distratora começam juntas até que o estímulo auditivo seja iniciado, quando ocorre o distanciamento das fixações para distratores. O cálculo da razão entre competidor e distratores foi significativo para o grupo de falantes nativos e para o grupo de falantes não nativos de inglês. O maior tempo de fixação ocular reflete a dificuldade em distinguir os pares de palavras terminados nas consoantes nasais /m/ e /n/.

Os resultados empírico-experimentais, portanto, corroboram a dificuldade do brasileiro em distinguir e produzir as consoantes nasais /m/ e /n/ em final de sílaba discutida em estudos anteriores (ZIMMER *et al.*, 2009; SILVEIRA, 2012; BECKER, 2014). Inclusive, a distinção não ocorreu com acurácia em alguns casos. Porém, o presente estudo avança em relação a estudos anteriores por apontar que tal dificuldade é passível de ser superada pelo uso linguístico oral mesmo em situação de não imersão geográfica em comunidades falantes da língua alvo.

A métrica de uso linguístico oral em segunda língua extraída do questionário de levantamento de usos linguísticos foi validada como previsora do percurso de competição lexêmica para o reconhecimento de palavras. Tal métrica se sobrepôs às demais medidas (amplitude lexical e acurácia de produção) realizadas para comparação. Portanto, poderá ser reutilizada para estudos futuros relacionados aos modelos baseados no uso para além das medidas de proficiência e tempo de imersão. Sobretudo, compreendemos que tais observações oferecem à comunidade científica que investiga a aquisição e processamento de L2, especificamente por bilíngues tardios e imersos em ambiente sociolinguístico onde a L2 não é socialmente dominante, uma observação que contribui para a validação externa de parte do questionário de uso utilizado em Valadares (2017) como preditor de comportamento linguístico de usuários de L2.

O maior uso linguístico oral em L2 em um contexto de imersão em país de língua materna apontado pelas respostas às questões de oralidade na etapa dois da coleta de dados se mostrou suficiente para a maior capacidade discriminatória de distribuições fonotáticas exclusivas da L2 revelada pelo

modelo de regressão linear. Esta observação vem ao encontro dos crescentes relatos anedóticos de aprendizagem de línguas adicionais como consequência do acesso a mídias eletrônicas destinadas primariamente ao lazer. Assim, este estudo reitera a significância do uso para o processamento linguístico bilíngue. Julgamos essa ponderação particularmente relevante para os estudos sobre aquisição de L2, pois ela desmistifica a noção amplamente difundida de que apenas a imersão em país de segunda língua é eficaz para proporcionar alto nível de desenvolvimento linguístico em uma segunda língua.

O ensino e aprendizagem de L2 na atualidade coloca como objetivo da aquisição fonológica a inteligibilidade, não o objetivo de indiferenciação com a pronúncia dos falantes nativos, o que talvez reflita a crescente aceitação de que essa aprendizagem resulte em estados de competência linguística que não devem nem podem parametrizar-se a partir do estado de competência monolíngue. Não obstante, ainda é um objetivo científico legítimo e necessário desvendar-se que fatores são constitutivos das diferenças que podem ser verificadas no processamento da linguagem por falantes não nativos e falantes nativos. As representações fonológicas e suas implementações fonéticas são seguramente aspectos em que essa diferenciação é bastante notória. Nossos resultados sugerem que a intensidade de uso seja um mecanismo central na aquisição de representações fonológicas, e esse mecanismo parece-nos interessante por ser capaz de explicar o impacto de outros fatores associados à aquisição dessas representações, tais como a idade e a própria imersão no ambiente da L2. Igualmente, nossos resultados sugerem que iniciativas de planejamento pedagógico que promovam ao máximo a integração das mídias de suporte ao discurso oral às trajetórias de aprendizagem podem ter impactos positivos no aprimoramento da capacidade de uso adequado de uma segunda língua por aprendizes.

Obviamente, caberá a estudos futuros a ampliação de contrastes fonêmicos entre a L1 e a L2 a serem investigados à luz da modulação do uso por nós observada, para que a hipótese considerada no presente estudo possa se provar robusta o suficiente para amplas generalizações sobre a aquisição e o processamento de segunda língua.

REFERÊNCIAS

ALLOPENA, Paul D.; MAGNUSON, James S.; TANENHAUS, Michael K. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439, 1998.

AVERY, Peter; EHRLICH, Susan. *Teaching American English Pronunciation*. Oxford: Oxford University Press, 1992.

BARR, Dale J. Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457-474, 2008.

BASSETTI, Benedetta. Orthographic input and second language phonology. In.: PISKE, Thorsten; YOUNGSCHOLTEN, Martha (ed.), *Input Matters in SLA*. Clevedon: Multilingual Matters, 191-206, 2008.

BECKER, Marcia Regina. Análise acústica das consoantes nasais bilabiais e alveolares do inglês. *Revista X*, vol. 1, 102-114, 2014.

BUNDGAARD-NIELSEN, Rikke L. ; BEST, Catherine T.; TYLER, Michael Douglas. Vocabulary size matters: The assimilation of L2 Australian English vowels to L1 Japanese categories. *Applied Psycholinguistics*, 32 , 51 – 67, 2011.

BUNDGAARD-NIELSEN, Rikke. L. ; BEST, Catherine T.; TYLER, Michael Douglas. Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 33 , 433– 461, 2011b.

BURTON, Martha W.; JONGMAN, Allard, K.; SERENO, Joan A. Phonological and orthographic priming effects in auditory and visual word recognition. *34th annual meeting of the Psychonomic Society*, Washington, DC. Nov. 1993.

BYBEE, Joan. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.

BYBEE, Joan. *Language, Usage and Cognition*. Cambridge: Cambridge University Press, 2010.

DAHAN, Delphine; MAGNUSON, James S.; TANENHAUS, Michael K. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317–367, 2001.

DAHAN, Delphine.; TANENHAUS, Michael K. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, 12, 453-459, 2005.

EWEN, Collin J.; HULST, Harry van der. *The phonological structure of words: an introduction*. Cambridge: Cambridge University Press, 2001.

FERNÁNDEZ, Eva M.; CAIRNS, Helen Smith. *Fundamentals of Psycholinguistics*. Malden: Wiley-Blackwell, 2010.

GONTOW, Cris; MARCELINO, Marcello; GODOY, Sonia M. B. *English Pronunciation for Brazilians: The sounds of American English*. São Paulo: Disal, 2006.

GROSJEAN, François. The Complementarity Principle and its impact on processing, acquisition, and dominance. In.: SILVA-CORVALÁN, Carmen.; TREFFERS-DALLER, Jeanine. (ed.) *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press, 2016.

GUAN, Guiyun; HU, Axu; GUO, Lei; YU, Hongzhi. An Eye-tracking Technique - Visual-World Paradigm, Advances in Social Science, Education and Humanities Research. *Proceedings of 3rd International Conference on Economic Development and Education Management (ICEDEM 2019)*, vol. 347, 2019.

HEREDIA, Roberto R. Bilingual memory and hierarchical models: A case for language dominance. *Current Directions in Psychological Science*, 34-39, 1997.

HUETTIG, Falk; ALTMANN, Gerry T.M. The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In.: CARREIRAS, Manuel; CLIFTON, Charles E. (ed.) *The on-line study of sentence comprehension: Eyetracking, ERP and beyond*. New York: Psychology Press, 187-207, 2004.

HUETTIG, Falk; MCQUEEN, James M., The tug war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, 57, 460-482, mar 2007.

HULSTIJN, Jan H. The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, 15(2), 422-433, abr 2012.

JOHNSON, Keith. Speech perception without speaker normalization. In.: JOHNSON, Keith; MULLENNIX, John. (ed.) *Talker variability in speech processing*. San Diego: Academic Press, p. 145-165, 1997.

JOHNSON, Keith. Decisions and Mechanisms in Exemplar-based Phonology. *UC Berkeley Phonology Lab Annual Report*, 289-311, 2005.

KEMPEN, Gerard; HUIJBERS, Pieter. The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, 14:185-209, set 1983.

KEMPEN, Gerard; HOENKAMP, Edward. An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11:201–58, abr/ jun 1987.

KROLL, Judith F.; DUSSIAS, Paola E.; BOGULSKI, Cari A.; VALDES-KROFF; Jorge. Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In.: ROSS, Brian H. (ed.) *The Psychology of Learning and Motivation*. San Diego: Academic Press, vol. 56, 229–262, 2012.

LEVELT, Willem. J.M. *Speaking: From intention to articulation*. Cambridge: MIT Press, 1989.

LEVELT, Willem J.M.; ROELOFS, Ardi; MEYER, Antje S. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 22, 1–75, fev 1999.

LIVERSEDGE, Simon P.; FINDLAY, John M. Saccadic eye movements and cognition, *Trends in Cognitive Sciences*, 4 (1), jan. 2000.

MARIAN, Viorica, SHOOK, Anthony. The Cognitive Benefits of Being Bilingual. *Cerebrum*, 1-12, 2012. Disponível em: https://dana.org/cerebrum/2012/the_cognitive_benefits_of_being_bilingual/. Acesso em: 04 de set. de 2019.

MATTOSO CÂMARA, JR., Joaquim. *Estrutura da língua portuguesa*. Petrópolis: Vozes, ed. 28., 1998 [1970].

MCQUEEN, James M.; VIEBAHN, Malte C. Tracking recognition of spoken words by tracking looks to printed words, *The Quarterly Journal of Experimental Psychology*, 60:5, 661-671, maio 2007.

MIRMAN Daniel; DIXON, James A.; MAGNUSON, James S. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59: 475-494, 2008.

MIRMAN, Daniel.; MAGNUSON, James S. Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, 37(7), 1026-1039, 2009.

MUNRO, Murray J.; DERWING, Tracey M. Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58, 479-502, 2008.

NATION, I.S. Paul; BEGLAR, David. A vocabulary size test. *The Language Teacher*, 31(7), 9-13, 2007.

PIERREHUMBERT, Janet B. Exemplar dynamics: Word frequency, lenition, and contrast. In.: BYBEE, Joan; HOPPER, Paul (ed.). *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 137-157, 2001.

PIERREHUMBERT, Janet B. Probabilistic phonology: discrimination and robustness. In.: BOD, Rens; HAY, Jennifer (ed.). *Probabilistic linguistics*, MIT Press, 2003.

SALVERDA, Anne Pier.; DAHAN, Delphine; TANENHAUS, Michael K.; CROSSWHITE, Katherine; MASHAROV, Mikhail; MCDONOUGH, Joyce. Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, 105: 466-76, 2007.

SALVERDA, Anne Pier; TANENHAUS, Michael K. Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. 36: 1108-17, 2010.

SEIDENBERG, Mark S.; TANENHAUS, Michael K. Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 546–554, 1979.

SEKERINA, Irina A. Visual World Eye-tracking Paradigm, In.: BROOKS, Patricia J.; KEMPE, Vera (ed.) *Encyclopedia of Language Development*, Londres: Sage, 2014.

SHEN, Wei; QU, Qingqing; TONG, Xiuhong. Visual attention shift to printed words during spoken word recognition in Chinese: The role of phonological information, *Memory & Cognition*, 46, 642–654, mai 2018.

SILVEIRA, Rosane. L2 Production of English Word-Final Consonants: The Role of Orthography and Learner Profile Variables. *Trabalhos em Linguística Aplicada (UNICAMP)*, v. 51, p. 15-28, 2012.

TANENHAUS, Michael K.; MAGNUSON, James S.; DAHAN, Delphine.; CHAMBERS, Craig. Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29, 557–580, 2000.

TANENHAUS, Michael K.; SPIVEY-KNOWLTON, Michael J. Eyetracking. *Language and Cognitive Processes*, 11, 583–588, 1996.

TANENHAUS, Michael K.; SPIVEY-KNOWLTON Michael J.; EBERHARD, Kathleen; SEDIVY,

Julie C. Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634, jun 1995.

TANENHAUS, Michael K., TRUESWELL, John C. Eye Movements and Spoken Language Comprehension. In.: TRAXLER, Matthew J.; GERNSBACHER, Morton Ann. (ed.) *Handbook of Psycholinguistics*, 863-900, USA: Elsevier, ed.2, 2006.

VALADARES, Marcus Guilherme Pinto de Faria. *Padrões emergentes de dominância linguística em português e inglês: O impacto de práticas socioculturais de letramento (digital) na amplitude lexical de brasileiros falantes de língua inglesa como L2*. Belo Horizonte: Universidade Federal de Minas Gerais, 2017. (Tese de doutorado)

VIDAL, Karina. A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61, 219–258, 2011.

WEBER, Andrea; MELINGER, Alissa; TAPIA, Lourdes Lara. The mapping of phonetic information to lexical representations in Spanish: evidence from eye movements, *ICPhS XVI*, 1941-1944, aug 2007.

YEE, Eiling.; SEDIVY, Julie C. Eye movements to pictures reveal transiente semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32 (1), 1-14, 2006.

ZIEGLER, Johannes C.; FERRAND, Ludovic. Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, 5, 683–689, 1998.

ZIMMER, Márcia Cristina. A.; SILVEIRA, Rosane; ALVES, Ubiratã Kickhöfel. *Instruction for Brazilians: Bringing Theory and Practice Together*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2009.

APÊNDICE - Tabela de Dados Tabulados

Participante	Amplitude Lexical em L2	Produção Oral em L2	Perfil de Uso Linguístico em L2	Tempo de Fixação nas Palavras Alvo	Tempo de Fixação nas Palavras Competidoras	Diferença entre Fixação nas Palavras Alvo e Competidoras
A	48	22	8	51,7	38,05	13,65
B	37	14	9	44,16	43,81	0,35
C	69	26	14	43,59	40,07	3,52
D	75	25	14	56,56	31,54	25,02
E	74	25	10	40,81	33,46	7,35
F	54	24	4	50,26	39,37	10,89
G	61	26	15	52,46	37,57	14,89
H	63	24	16	56,34	33,16	23,18
I	57	17	13	44,98	44,14	0,84
J	73	24	15	66,89	19,45	47,44
K	89	28	17	52,46	30,82	21,64
L	72	22	13	45,63	43,4	2,23
M	88	28	16	55,38	29,41	25,97
N	74	26	16	60,01	26,14	33,87
O	45	19	11	48,87	39,47	9,4
P	73	15	12	52,41	37,46	14,95
Q	67	26	15	53,17	36,69	16,48
R	83	21	11	48,23	40,42	7,81
S	43	25	13	49,82	41,03	8,79
T	48	22	9	44,6	40,47	4,13
U	32	27	7	50,6	37,78	12,82
V	75	27	11	40,5	42,65	-2,15
X	47	28	14	48,87	40,57	8,3
Y	80	24	15	59,59	27,7	31,89
Z	83	28	12	54,33	32,09	22,24
A1	75	24	15	53,83	31,42	22,41
B1	89	28	15	63,04	26,55	36,49
C1	75	30	15	54,55	29,69	24,86

VARIÁVEIS LEXICAIS E ORTOGRÁFICAS NO ACESSO LEXICAL DAS PALAVRAS DO PORTUGUÊS BRASILEIRO

LEXICAL AND ORTHOGRAPHIC VARIABLES IN THE LEXICAL ACCESS OF WORDS IN BRAZILIAN PORTUGUESE

Gustavo Lopez Estivalet¹

RESUMO

Quais variáveis influenciam o acesso lexical no português brasileiro? O presente trabalho teve como objetivo analisar a influência das principais variáveis lexicais e ortográficas no acesso lexical: a. frequência lexical, b. vizinhança ortográfica, c. categoria gramatical, d. número de letras e e. número de sílabas. Modelos de leitura de via dupla postulam diferentes rotas para o processamento lexical e ortográfico, nos quais essas variáveis influenciam diferentemente as etapas do processamento das palavras. Portanto, para uma melhor compreensão do léxico mental, as influências dessas variáveis devem ser analisadas no português brasileiro. Para tanto, aplicou-se um experimento psicolinguístico com tarefa de decisão lexical na modalidade visual entre palavras e pseudopalavras selecionadas e criadas a partir do Léxico do Português Brasileiro. Os resultados apontaram as seguintes diferenças significativas: palavras foram reconhecidas mais rapidamente do que pseudopalavras; palavras de alta frequência foram reconhecidas mais rapidamente do que palavras de média frequência, que por sua vez, foram reconhecidas mais rapidamente do que palavras de baixa frequência; palavras com alta vizinhança ortográfica foram reconhecidas mais rapidamente do que palavras de baixa vizinhança. Esses resultados sugerem que essas variáveis influenciam o processamento das palavras no acesso lexical. Ainda, os resultados das variáveis número de letras e número de sílabas indicam efeitos graduais no reconhecimento das palavras. Assim, a presente investigação contribui para uma melhor compreensão dos processos de leitura durante o acesso lexical, assim como valida a seleção, criação e controle de estímulos com o uso de corpora na pesquisa em psicolinguística experimental no português brasileiro.

Palavras-chave: Psicolinguística; Acesso Lexical; Reconhecimento de Palavras; Léxico Mental; Decisão Lexical.

ABSTRACT

Which variables influence lexical access in Brazilian Portuguese? The present work aimed to analyze the influence of the main lexical and orthographic variables on lexical access: a. lexical frequency, b. orthographic neighborhood, c. grammatical category, d. number of letters, and e. number of syllables. Dual-route models of reading postulate different routes for lexical and orthographic processing, in which these variables differently influence the stages of word processing. Therefore, for a better understanding of the mental lexicon, the influences of these variables should be analyzed in Brazilian Portuguese. To this end, it was applied a psycholinguistic experiment with a lexical decision task in visual modality between words and pseudowords selected and created from the Brazilian Portuguese Lexicon. The results showed the following significant differences: words were recognized faster than pseudowords; high-frequency words were recognized faster than medium-frequency words, which were recognized faster than low-frequency words; high orthographic neighborhood words were recognized faster than low orthographic neighborhood words. These results suggest that these variables influence word processing in lexical access. Furthermore, the results of the variables number of letters and number of syllables indicate gradual effects on word recognition. Thus, the present investigation contributes to a better understanding of reading processes during lexical access, as well as, validate the selection, creation, and control of stimuli with the use of corpora in experimental psycholinguistic research in Brazilian Portuguese.

Keywords: Psycholinguistics; Lexical Access; Word Recognition; Mental Lexicon; Lexical Decision.

¹ Departamento de Letras Estrangeiras Modernas, Universidade Federal da Paraíba (UFPB) - Laboratório de Processamento Linguístico (LAPROL). Email: gustavoestivalet@hotmail.com.

Introdução

Os avanços das pesquisas em psicolinguística têm demonstrado que uma série de variáveis influencia diferentemente o processamento lexical e ortográfico durante o reconhecimento das palavras e seu acesso lexical. Trabalhos com metodologias *online* em palavras isoladas do inglês, utilizando tarefas simples de decisão lexical, têm apontado principalmente as seguintes variáveis: frequência lexical (MONSELL; DOYLE; HAGGARD, 1989), vizinhança ortográfica (ANDREWS, 1997) e número de letras (NEW *et al.*, 2006).

Contudo, a maior parte desses trabalhos tem sido realizada em inglês, holandês, francês, alemão e espanhol (MARIAN *et al.*, 2012), mas poucas pesquisas são feitas no português, e, menos ainda, no português brasileiro (PB). Possivelmente, isso se deve à escassez de recursos metodológicos disponíveis e à falta de publicações em psicolinguística experimental relacionadas ao léxico mental e acesso lexical no PB (ESTIVALET; MEUNIER, 2017). Sendo assim, faz-se necessário o aprofundamento da investigação empírica dessas variáveis no reconhecimento de palavras no PB.

Destaca-se que o efeito de frequência é um dos principais fenômenos encontrados, replicados e discutidos nos estudos empíricos do léxico mental. Igualmente, os efeitos de tamanho de palavras (número de letras e número de sílabas), assim como vizinhança ortográfica, são efeitos recorrentes em psicolinguística, que por sua vez, têm sido explicados através de diferentes modelos empíricos (TAFT, 1991). Enquanto modelos seriais predizem esses efeitos independentes em diferentes etapas do processamento lexical (BESNER; JOHNSTON, 1992), modelos interativos predizem a saliência desses efeitos a partir das interações dos diferentes nós de processamento (RUMELHART; MCCLELLAND, 1982).

Monsell *et al.*, (1989) apresentaram um estudo com experimentos e simulações, evidenciando que os efeitos de frequência podem acontecer em diferentes etapas da tarefa de decisão lexical: i. decodificação ortográfica, ii. identificação lexical e iii. processo de decisão. Taft (1991) revisou uma série de experimentos em inglês com manipulação de frequências de lexema e lema. O autor propôs um modelo de decomposição obrigatória em que o acesso lexical é guiado por ambas frequências, sendo a frequência de lexema interpretada como a frequência que os radicais se combinam com os sufixos. O modelo de decomposição obrigatória é composto por três etapas: i. decomposição, ii. ativação e iii. recombinação. Na primeira etapa, palavras morfologicamente complexas são decompostas a partir de evidências estruturais e de saliência; em seguida, os morfemas são procurados no léxico mental e seus traços lexicais, semânticos e morfossintáticos são ativados; enfim, os morfemas são

recombinados para verificação da existência da palavra. Portanto, efeitos de frequência de lema acontecem na primeira e na segunda etapa durante a decomposição, procura e ativação dos morfemas no léxico mental, enquanto os efeitos de frequência de lexema acontecem somente na terceira etapa durante a verificação da recombinação dos morfemas e existência da palavra.

New *et al.* (2006) mostraram que os efeitos de tamanho de palavras também são pertinentes para o acesso lexical, os autores demonstraram que os tempos de reconhecimento lexical de palavras do inglês obedecem uma curva em U: palavras de 3 e 4 letras apresentam tempos maiores, palavras de 5 a 8 letras apresentam os menores tempos (sendo 7 letras o menor tempo) e palavras de 9 a 13 apresentam tempos maiores. Andrews (1997) demonstrou através de uma série de experimentos a estreita relação entre vizinhança ortográfica e frequência lexical. Palavras com muitos vizinhos apresentam tempos de reconhecimento mais rápidos do que palavras com poucos vizinhos, pois apesar da concorrência lexical, a ativação de outras palavras semelhantes é um forte indício da existência da palavra alvo.

Com o objetivo de contemplar estes diferentes achados empíricos sobre o processamento de palavras e acesso lexical, Besner e Johnson (1992) propuseram um modelo complexo de reconhecimento de palavras de via dupla. Na via lexical, as palavras são reconhecidas diretamente através da familiaridade visual; já na via de análise visual, as palavras são processadas através de diferentes componentes: i. análise ortográfica, ii. silabação, iii. similaridade ortográfica/fonológica, iv. frequência lexical e v. ativação de traços morfossintáticos.

Neste sentido, o presente trabalho teve como objetivo principal investigar a influência dessas três variáveis (frequência lexical, vizinhança ortográfica e categoria gramatical) no processamento de palavras do PB. Como objetivos secundários, exploraram-se outras duas variáveis pertinentes: número de sílabas (MEHLER *et al.*, 1981) e categoria gramatical (CARAMAZZA; HILLIS, 1991). Portanto, esta pesquisa justifica-se a partir de dois aspectos: i. investigação das variáveis lexicais e ortográficas no processamento das palavras do PB, e, ii. verificação das normas de frequência e vizinhança ortográfica do *corpus* Léxico do Português Brasileiro (LexPorBR)² (ESTIVALET; MEUNIER, 2015).

2 O Léxico do Português Brasileiro (LexPorBR) é um léxico baseado em palavras construído com o objetivo de auxiliar e guiar a seleção, o controle e a criação de palavras e pseudopalavras do português brasileiro para o desenvolvimento de experimentos psicolinguísticos. Ele foi elaborado a partir do corpus do NILC de textos escritos e contém mais de 42 milhões de palavras *token*, 215 mil palavras *type* e 21 categorias de informações lexicais e gramaticais (ESTIVALET; MEUNIER, 2015; 2017). O LexPorBR pode ser acessado, consultado e baixado gratuitamente no site: <http://www.lexicodoportugues.com>.

A hipótese de base é que somente as duas primeiras variáveis citadas acima (i.e., frequência lexical e vizinhança ortográfica) influenciam o processamento e reconhecimento de palavras escritas do PB (TAFT, 1991). Diferentemente, a hipótese da presente investigação é que todas as cinco diferentes variáveis lexicais e ortográficas influenciam diferentemente o processamento e reconhecimento das palavras (BAAYEN, 2008). A hipótese alternativa é que interações entre as informações de frequência, ortográficas e semânticas definem o processamento e reconhecimento das palavras (RUMELHART; MCCLELLAND, 1982).

Para tanto, foi aplicado um experimento psicolinguístico *online* com uma tarefa de decisão lexical com 720 palavras selecionadas a partir do *corpus* LexPorBR (ESTIVALET; MEUNIER, 2015). Os resultados apresentaram diferenças significativas entre palavras de alta, média e baixa frequência (MONSELL; DOYLE; HAGGARD, 1989) e palavras com alta e baixa vizinhança ortográfica (ANDREWS, 1997). Ainda, foram encontradas diferenças significativas em relação ao número de letras (NEW et al., 2006) e número de sílabas (MEHLER *et al.*, 1981). Esses resultados são discutidos à luz de modelos psicolinguísticos de reconhecimento das palavras, sugerindo evidências a favor de modelos que consideram a complexidade e influência dessas diferentes variáveis no sistema de processamento e reconhecimento das palavras no acesso lexical (BESNER; JOHNSTON, 1992).

2. Metodologia de pesquisa

2.1 Participantes

Foram testados 24 participantes (12 mulheres) falantes do PB como língua materna entre 18 e 42 anos (idade média 27,1 anos). Todos participantes eram estudantes de cursos universitários, destros, possuíam audição normal, visão normal ou corrigida por óculos/lentes de contato e não possuíam histórico de problemas cognitivos ou de linguagem. Os participantes não sabiam os objetivos do estudo e assinaram um termo de consentimento para participação no experimento como voluntários. O experimento foi conduzido de acordo com a Declaração de Helsinki (WORLD MEDICAL ASSOCIATION, 2013), tendo o protocolo aprovado pelo Comitê de Ética de Pesquisas com Seres Humanos IRB: 00009118.

2.2 Materiais

Os participantes realizaram um experimento com uma tarefa de decisão lexical entre palavras e pseudopalavras na modalidade visual. Foram investigadas três variáveis fatoriais: a. frequência lexical

(alta, média e baixa), b. vizinhança ortográfica (alta e baixa), c. categoria gramatical (substantivos e verbos); e, duas variáveis numéricas: d. número de letras (3-15) e e. número de sílabas (1-6). A Tabela 1 apresenta exemplos de estímulos experimentais.

Tabela 1: Exemplos de palavras nas condições experimentais..

Cat. Gram.	Viz./Freq.	Alta frequência	Média frequência	Baixa frequência
Substantivos	Alta	campanha	barcos	abrigo
	Baixa	juros	beijos	fico
Verbos	Alta	acabei	somaram	deram
	Baixa	ocupa	exercida	pede

Fonte: elaboração própria.

Foram selecionadas 720 palavras do PB como estímulos experimentais, 360 substantivos e 360 verbos. Dentro de cada categoria gramatical, 180 palavras foram de alta vizinhança ortográfica (>4) e 180 palavras de baixa vizinhança ortográfica (<4), de acordo com o N de Coltheart (COLTHEART *et al.*, 1977; YARKONI; BALOTA; YAP, 2008). Dentro de cada conjunto de vizinhança ortográfica, 60 palavras foram de alta frequência (>4), 60 palavras de média frequência (>3 e <4) e 60 palavras de baixa frequência (<3), de acordo com a escala *zipf* (VAN HEUVEN *et al.*, 2014). As variáveis investigadas (frequência lexical, vizinhança ortográfica, categoria gramatical, número de letras e número de sílabas) foram retiradas do léxico LexPorBR (ESTIVALET; MEUNIER, 2015). Destaca-se que essas métricas foram contabilizadas a partir do Corpus do NILC³ (Núcleo Interinstitucional de Linguística Computacional) contendo mais de 42 milhões de palavras *token* e 215 mil palavras *type* (PINHEIRO; ALUÍSIO, 2003). A escala *zipf* é uma linearização da distribuição das frequências lexicais que pode ser comparada entre diferentes *corpora* (VAN HEUVEN *et al.*, 2014); o número de letras foi calculado a partir do número de caracteres das palavras; o número de sílabas foi calculado a partir da aplicação do silabador *Petrus* do NILC (MARQUIAFAVEL; BOKAN; ZAVAGLIA, 2014); e as normas de vizinhança ortográfica foram calculadas a partir do pacote 'vwr' (KEULEERS, 2013). Todos estímulos experimentais foram selecionados e controlados em i. frequência lexical, ii. número de letras, iii. número de sílabas, iv. vizinhança ortográfica e v. OLD20, conforme a Tabela 2. As 720 pseudopalavras foram criadas a partir do motor de geração de pseudopalavras do *corpus* LexPorBR (ESTIVALET; MEUNIER, 2017).

³ O corpus do NILC (Núcleo Interinstitucional de Linguística Computacional) é formado por textos escritos, contendo diversos gêneros textuais, conforme a seguinte distribuição do número de palavras *token*: texto didático 182.679, enciclopédia 286.317, ensaio 2.191.189, texto jornalístico 30.179.462, texto legal 1.067.575, literário 923.133, revista 153.78 (PINHEIRO; ALUÍSIO, 2003). O corpus do NILC pode ser acessado, consultado e baixado nos sites: <https://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>.

Tabela 2: Médias das características lexicais das palavras nas condições experimentais.

Cat.	Frequência	Vizinhança	Frequência	Letras	Sílabas	Vizinhos	OLD20
Substantivos	Alta	Alta	4,61	6,97	2,53	7,30	1,49
		Baixa	4,28	7,53	2,97	1,33	2,15
	Média	Alta	3,52	6,07	2,43	6,50	1,57
		Baixa	3,45	8,60	3,53	0,73	2,56
	Baixa	Alta	2,76	6,70	2,40	7,23	1,52
		Baixa	2,90	8,30	3,43	0,67	2,62
Verbos	Alta	Alta	4,39	6,03	2,50	6,27	1,51
		Baixa	4,23	7,53	3,10	1,63	1,86
	Média	Alta	3,43	6,17	2,57	6,63	1,54
		Baixa	3,45	8,37	3,70	1,53	1,90
	Baixa	Alta	2,83	6,43	2,70	5,53	1,61
		Baixa	2,88	8,60	3,70	1,47	2,02

Fonte: elaboração própria.

Com o objetivo de se testar um grande número de (pseudo)palavras do PB (720 palavras e 720 pseudopalavras), os 1.440 estímulos foram divididos em oito blocos de 180 palavras. A partir desses oito blocos, construíram-se 12 listas experimentais contendo quatro blocos cada uma (i.e., 720 estímulos), seguindo-se o contrabalanceamento de quadrado-latino (BALOTA *et al.*, 2007). As listas possuíam os seguintes critérios: i. não houve duas palavras seguidas começando com a mesma letra e ii. no máximo três palavras ou pseudopalavras com as mesmas condições experimentais em sequência. Esses critérios foram estabelecidos com o objetivo de minimizarem-se possíveis efeitos de *priming* ortográfico/fonológico, assim como das características das variáveis pesquisadas. O experimento final contou com 720 estímulos de teste e 10 estímulos de prática, houve um intervalo de um minuto no meio do experimento e o tempo médio de realização do experimento foi de 18 minutos.

2.3 Procedimentos

O experimento foi construído e aplicado através do programa E-Prime® 2.0 Professional (Psychology Software Tools, Inc., Sharpsburg, PA, USA) (SCHNEIDER; ESCHMAN; ZUCCOLOTTO, 2012). Os participantes foram testados individualmente em local silencioso. Cada estímulo foi apresentado na seguinte sequência: primeiro, um ponto de fixação era apresentado por 500 ms; segundo, a palavra ou pseudopalavra era apresentada por 2.000 ms ou até a resposta do participante; terceiro, uma tela preta era apresentada por 500 ms; enfim, a mesma sequência recomeçava com a apresentação de um novo estímulo. Os estímulos foram apresentados no centro da tela LCD 15" de um computador portátil em letras minúsculas e brancas sobre um fundo preto,

em fonte Courier New e tamanho 18 pt. A mensuração dos tempos de reação (RT) começava com a apresentação da palavra ou pseudopalavra e terminava quando o participante realizava sua resposta. Os participantes utilizavam uma tecla verde do teclado do computador para palavras existentes e uma tecla vermelha do teclado do computador para pseudopalavras inexistentes. Eles foram orientados a realizarem suas respostas o mais corretamente e rapidamente possível.

3. Resultados e discussão

Encontrou-se uma diferença significativa entre os RT de palavras 679(234) ms e pseudopalavras 844(338) ms, $t(46) = 4,72$, $p < 0,001$ (teste-t de Student, distribuição bicaudal, amostras pareadas), assim como uma diferença significativa entre as taxas de erro de palavras 1,48% e pseudopalavras 2,51%, $t(46) = 2,33$, $p < 0,05$ (teste-t de Student, distribuição bicaudal, amostras pareadas). Esse efeito é recorrente em tarefas de decisão lexical, pseudopalavras são descartadas mais lentamente que palavras são reconhecidas devido ao tempo de procura do estímulo no léxico mental. Ainda, pseudopalavras apresentaram maior taxa de erro porque eventualmente são classificadas como palavras potenciais da língua (DUYCK *et al.*, 2004). Portanto, destaca-se que as pseudopalavras geradas pelo LexPorBR efetivamente respeitaram as estruturas ortotáticas do PB, resultando em confusão com palavras potenciais (ESTIVALET; MEUNIER, 2017). A seguir, somente as palavras existentes foram analisadas.

RT abaixo de 300 ms e acima de 1.800 ms foram considerados fora da tarefa e descartados (0,46%), uma palavra experimental (i.e., emana) foi removida porque apresentou taxa de erro superior a 50% (0,06%) e respostas incorretas foram removidas (2,96%), no total 3,47% dos dados foram descartados para análise dos RT (BAAYEN, 2008). Essa limpeza assegura uma melhor distribuição normal dos dados para análise dos mesmos. As médias dos RT, desvios padrão e taxas de erro são apresentados na Tabela 3.

Tabela 3: Médias dos RT, desvios padrão e de erro nas condições experimentais fatoriais.

Cat. Gram.	Freq. Viz.	Alta		Média		Baixa	
		RT(ms)	Erro(%)	RT(ms)	Erro(%)	RT(ms)	Erro(%)
Substantivos	Alta	586(152)	0,06	649(185)	0,23	696(211)	0,60
	Baixa	629(173)	0,02	729(232)	0,13	761(243)	0,20
Verbos	Alta	614(157)	0,13	649(191)	0,35	704(220)	0,80
	Baixa	615(164)	0,09	665(197)	0,12	724(233)	0,24

Fonte: elaboração própria.

As variáveis fatoriais foram avaliadas através de uma análise de variância (ANOVA, F1:

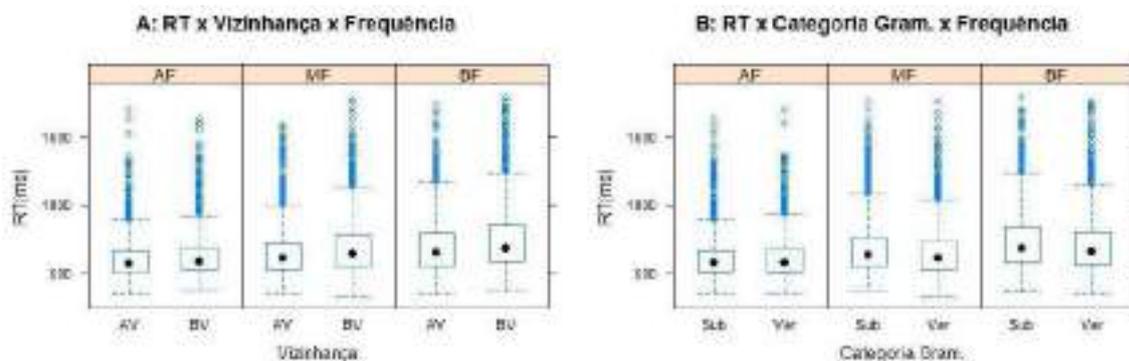
participantes) contendo RT como variável dependente; a. frequência lexical (alta, média e baixa), b. vizinhança ortográfica (alta e baixa) e c. categoria gramatical (substantivos e verbos) como variáveis independentes fixas; e participantes como variável aleatória (BAAYEN, 2008). Foram encontrados os seguintes efeitos principais significativos: frequência lexical $F(2,23) = 298,07, p < 0,001$, vizinhança ortográfica $F(1,23) = 110,02, p < 0,001$ e categoria gramatical $F(1,23) = 6,73, p < 0,05$, e, os seguintes efeitos de interação significativos: frequência lexical e categoria gramatical $F(2,23) = 10,76, p < 0,001$ e vizinhança e categoria gramatical $F(1,23) = 49,28, p < 0,001$, conforme a Figura 1.

Corroborando o efeito de frequência, a presente tarefa de decisão lexical em PB encontrou diferenças significativas entre o reconhecimento de palavras de alta, média e baixa frequência lexical (MONSELL; DOYLE; HAGGARD, 1989). Palavras de alta frequência foram reconhecidas 62 ms mais rapidamente que palavras de média frequência $t(46) = 2,75, p < 0,05$ e palavras de média frequência foram reconhecidas 49 ms mais rapidamente que palavras de baixa frequência $t(46) = 2,32, p < 0,05$ (testes-t de Student pareados com correção de Bonferroni). Este resultado vai de encontro a teorias que propõem a organização do léxico mental em função das frequências lexicais.

Em seguida, encontrou-se o efeito de vizinhança ortográfica, em que palavras com alta vizinhança foram reconhecidas 39 ms mais rápido que palavras com baixa vizinhança. Este resultado facilitador em palavras com mais vizinhos ortográficos está de acordo com uma série de estudos apresentados por Andrews (1997), palavras com muitos vizinhos ortográficos ativam simultaneamente uma série de outras palavras e são rapidamente reconhecidas. Não houve efeito de interação entre frequência lexical e vizinhança ortográfica, sugerindo que esses efeitos são independentes (TAFT, 1991), conforme pode ser verificado na Tabela 3 e na Figura 1A. Observa-se que o efeito de vizinhança ortográfica ocorre principalmente em palavras pequenas, com até cinco letras, pois palavras grandes desassociam-se facilmente em função de sua complexidade ortográfica (MARIAN *et al.*, 2012).

Logo após, verificou-se que verbos são reconhecidos apenas 13 ms mais rapidamente que substantivos, esta diferença se deve às características do processamento morfológico entre substantivos derivados e verbos flexionados (CARAMAZZA; HILLIS, 1991). Apesar de os verbos apresentarem mais complexidade nos traços morfossintáticos, substantivos derivados apresentam maior complexidade estrutural e semântica durante seu processamento e reconhecimento. Mais importante que o efeito principal, houve uma interação entre frequência lexical e categoria gramatical, sugerindo que essas categorias gramaticais possuem um comportamento em palavras de alta frequência e outro comportamento em palavras de média e baixa frequência, conforme pode ser verificado na Figura 1B.

Figura 1: Variáveis fatoriais: a. frequência lexical, b. vizinhança ortográfica e c. categoria gramatical. AF: alta frequência, MF: média frequência, BF: baixa frequência, AV: alta vizinhança, BV: baixa vizinhança, Sub: substantivos, Ver: verbos.



Fonte: elaboração própria.

Esses resultados sugerem que os verbos são reconhecidos mais rapidamente que os substantivos somente em média frequência $t(46) = 2,24$, $p < 0,05$ e baixa frequência $t(46) = 2,83$, $p < 0,05$, mas não possuem diferenças em alta frequência, sugerindo um sistema de processamento lexical de duas vias (BESNER; JOHNSTON, 1992). Enfim, também houve um efeito de interação entre vizinhança ortográfica e categoria gramatical, justificado através do fato que substantivos possuem naturalmente mais vizinhos ortográficos do que verbos no PB (ESTIVALET; MEUNIER, 2015).

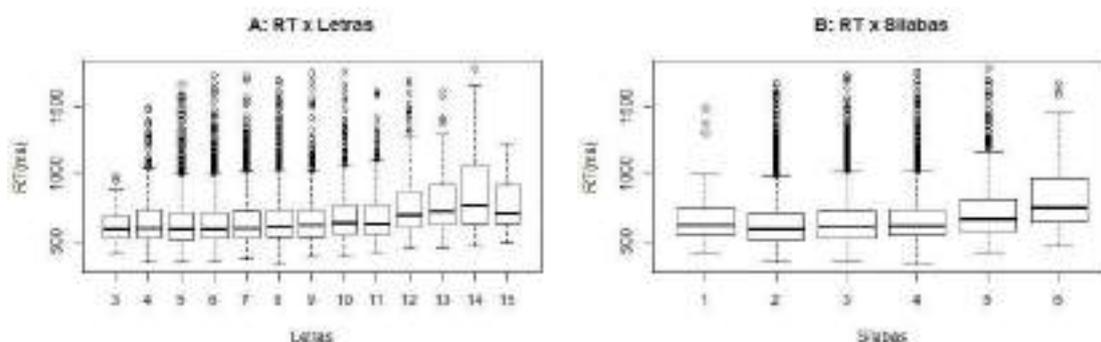
Na sequência, uma regressão linear múltipla foi calculada contendo os RT como variável dependente, e, d. número de letras (3-15) e e. número de sílabas (1-6) como variáveis independentes numéricas. Foi encontrado um efeito principal significativo $F(2,716) = 50,77$, $p < 0,001$ (número de letras $t(716) = 2,36$, $p < 0,05$; número de sílabas $t(716) = 2,51$, $p < 0,05$), com $R^2 = 0,12$, de acordo com a Figura 2.

Conforme encontrado em New *et al.* (2006), os resultados do presente estudo também apresentaram uma curva em U dos RT em relação ao número de letras, palavras de 3-6 letras possuem RT decrescentes e palavras de 7-15 letras possuem RT crescentes em função do número de letras, conforme a Figura 2A. Este efeito deve ser explicado em função da alta vizinhança ortográfica, assim como da maior homonímia e polissemia que palavras pequenas apresentam (MARIAN *et al.*, 2012). Importante, verifica-se que até nove letras, o reconhecimento das palavras em função do número de letras possui um comportamento semelhante, apresentando maiores diferenças e variâncias de 9-15 letras. Este efeito pode ser justificado devido à necessidade de mais de uma sacada ocular para a leitura de palavras maiores de nove letras (BESNER; JOHNSTON, 1992) e determinação da sílaba tônica.

Finalmente, da mesma forma que o comportamento do número de letras, os RT em relação ao número de sílabas também apresentou uma curva em U, com RT decrescentes para palavras de 1-2 sílabas e RT crescentes para palavras de 3-6 sílabas (MEHLER *et al.*, 1981). A mesma explicação dada acima em relação ao número de letras pode ser utilizada para o comportamento dos RT em função do número de sílabas. É necessário salientar que o processamento silábico inclui também um componente fonológico associado ao pé prosódico e sílaba tônica, logo, apresentando diferenças de RT ligados ao processamento temporal de leitura das palavras (BALOTA *et al.*, 2007).

Em geral, os resultados apresentados acima corroboram os efeitos psicolinguísticos de frequência lexical, vizinhança ortográfica e número de letras (TAFT, 1991), assim como a presente hipótese em relação aos efeitos de categoria gramatical e número de sílabas.

Figura 2: Variáveis numéricas d. número de letras e e. número de sílabas.



Fonte: elaboração própria.

Considerações finais

O principal objetivo do presente trabalho foi verificar a influência das três principais variáveis lexicais e ortográficas, frequência lexical, vizinhança ortográfica e número de letras (TAFT, 1991), no processamento de palavras isoladas do PB. Além dessas variáveis, também foram verificadas outras duas variáveis de interesse, categoria gramatical e número de sílabas, através de um experimento psicolinguístico com uma tarefa de decisão lexical em um grande número de palavras do PB (BALOTA *et al.*, 2007). Os resultados apresentaram efeitos principais significativos assim como algumas interações entre essas variáveis.

O efeito de frequência lexical tem sido discutido em diversos modelos de reconhecimento de palavras (MONSELL; DOYLE; HAGGARD, 1989). Modelos associativos de via única têm argumentado que a frequência lexical é a principal variável preditiva do resultado de interações

entre a forma e o significado das palavras (RUMELHART; MCCLELLAND, 1982). Contudo, estes modelos têm enfrentado dificuldades para explicar os diferentes efeitos de vizinhança ortográfica que emergem a partir da manipulação desta variável, assim como as diferenças entre o processamento de diferentes categorias gramaticais. O presente estudo sugere que os efeitos de vizinhança ortográfica estão presentes de forma independente da frequência lexical (MARIAN *et al.*, 2012). Ainda, a presente pesquisa encontrou diferenças no processamento e reconhecimento de palavras de diferentes categorias gramaticais, que por sua vez podem ser explicadas através do processamento dos traços morfossintáticos presentes nas diferentes categorias gramaticais, tais como substantivos derivados e verbos flexionados (CARAMAZZA; HILLIS, 1991).

Nesse sentido, o presente estudo contribui com evidências a favor de modelos de processamento de duas vias para o reconhecimento de palavras (BESNER; JOHNSTON, 1992). Enquanto algumas palavras são reconhecidas diretamente pela sua forma inteira no léxico mental, outras palavras são processadas através de seus constituintes linguísticos discretos. De uma forma geral, palavras de alta frequência são processadas pela via direta lexical enquanto palavras de média e baixa frequência são reconhecidas através do processamento ortográfico e fonológico. Ainda, substantivos derivados são acessados através da via direta no léxico mental e verbos flexionados são processados pelo componente morfossintático decomposicional.

Sendo assim, conforme proposto no modelo serial de Besner e Johnston (1992), os efeitos de frequência, vizinhança ortográfica, tamanho das palavras e categoria gramatical das palavras ocorrem em diferentes etapas do processamento lexical. Primeiramente, ocorrem os efeitos de tamanho de palavra em relação ao número de letras (unidades simbólicas a serem decodificados) e ao número de sílabas (unidades fonológicas a serem processadas); em seguida, ocorrem os efeitos de vizinhança ortográfica/fonológica em que outras palavras são ativadas, gerando competição lexical; após, ocorrem os efeitos de frequência de lexema e lema; enfim, ocorre o efeito de categoria gramatical com a ativação dos traços morfossintáticos.

Finalmente, o presente estudo também confirmou que as normas de frequência, vizinhança ortográfica, categoria gramatical, número de letras e número de sílabas estabelecidas pelo LexPorBR são pertinentes e devem ser observadas para a seleção, manipulação e controle de estímulos psicolinguísticos do PB (ESTIVALET; MEUNIER, 2017). Enfim, espera-se que novas pesquisas do processamento e reconhecimento de palavras venham enriquecer a compreensão deste e outros fenômenos, assim como colaborem para o desenvolvimento da psicolinguística experimental no PB.

REFERÊNCIAS

- ANDREWS, Sally. The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, v. 4, n. 4, p. 439–461, dez. 1997.
- BAAYEN, Rolf Harald. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press, 2008.
- BALOTA, David *et al.* The English Lexicon Project. *Behavior Research Methods*, v. 39, n. 3, p. 445–459, ago. 2007.
- BESNER, Derek; JOHNSTON, James. Reading and the mental lexicon: on the uptake of visual information. In: MARSLEN-WILSON, William (ed.). *Lexical Representation and Process*. Cambridge/MA: The MIT Press, p. 291–316, 1992.
- CARAMAZZA, Alfonso; HILLIS, Argye. Lexical organization of nouns and verbs in the brain. *Nature*, v. 349, n. 6312, p. 788–790, fev. 1991.
- COLTHEART, Max *et al.* Access to the internal lexicon. In: DORNIC, Stan (ed.). *Attention and Performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates, p. 535–555, 1977.
- DUYCK, Wouter *et al.* WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, v. 36, n. 3, p. 488–499, ago. 2004.
- ESTIVALET, Gustavo Lopez; MEUNIER, Fanny. The Brazilian Portuguese Lexicon: An Instrument for Psycholinguistic Research. *PLOS ONE*, v. 10, n. 12, p. e0144016, dez. 2015.
- ESTIVALET, Gustavo Lopez; MEUNIER, Fanny. Corpus psicolinguístico Léxico do Português Brasileiro. *Revista SOLETRAS*, v. 33, n. 1, p. 212–229, jul. 2017.
- KEULEERS, Emmanuel. vwr: Useful functions for visual word recognition research. *R Package*, p. 19, 2013.

MARIAN, Viorica *et al.* CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, v. 7, n. 8, p. e43230, ago. 2012.

MARQUIAFÁVEL, Vanessa; BOKAN, Alessandro; ZAVAGLIA, Cláudia. PETRUS: A rule-based grapheme-to-phoneme converter for Brazilian Portuguese. In: BAPTISTA, Jorge *et al.* (ed.). *PROPOR 2014, LNAI 8776*. Heidelberg, Germany: Springer, 2014.

MEHLER, Jacques *et al.* The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, v. 20, n. 3, p. 298–305, jun. 1981.

MONSELL, Stephen; DOYLE, Mark; HAGGARD, Patrick. Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, v. 118, n. 1, p. 43–71, 1989.

NEW, Boris *et al.* Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, v. 13, n. 1, p. 45–52, fev. 2006.

PINHEIRO, Gisele; ALUÍSIO, Sandra. *Corpus NILC: descrição e análise crítica com vistas ao projeto Lacio - WebSérie de Relatórios do Núcleo Interinstitucional de Linguística Computacional NILC - ICMC - USP*. São Carlos, SP: Universidade Federal de São Carlos - UFSCar, 2003.

RUMELHART, David; MCCLELLAND, Jay. An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, v. 89, n. 1, p. 60–94, 1982.

SCHNEIDER, Walter; ESCHMAN, Amy; ZUCCOLOTTO, Anthony. *E-Prime 2.0*. Learning Research and Development Center: University of Pittsburgh, 2012.

TAFT, Marcus. *Reading and the Mental Lexicon*. East Sussex/UK: Lawrence Erlbaum Associates Ltd., 1991.

VAN HEUVEN, Walter. *et al.* SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, v. 67, n. 6, p. 1176–1190, jun.

2014.

WORLD MEDICAL ASSOCIATION. World Medical Association Declaration of Helsinki. *JAMA*, v. 310, n. 20, p. 2191, nov. 2013.

YARKONI, Tal; BALOTA, David; YAP, Melvin. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, v. 15, n. 5, p. 971–979, out. 2008.

O PROCESSAMENTO DE MÚLTIPLAS DEPENDÊNCIAS NÃO ADJACENTES EM ORAÇÕES RELATIVAS NO PB¹

PROCESSING MULTIPLE NONADJACENT DEPENDENCIES IN BP

Késsia Henrique²

Cristina Name³

RESUMO

Neste artigo, investigamos o processamento de dependências não adjacentes (DNAs) e suas diferentes configurações no português brasileiro (PB), por falantes adultos. DNAs são definidas como a co ocorrência de marcadores morfossintáticos e/ou fonológicos, com material interveniente, como em *A* *menin-a*, em que se observa a dependência entre o determinante e a marca de gênero feminino *-a* no nome, tendo o material interveniente *menin-* entre os itens dependentes. No PB, do ponto de vista estrutural, há ocorrências de DNAs do tipo *aninhadas* (A1 A2 B2 B1), em que os elementos codependentes A2 B2 “se aninham” entre A1 B1, como em *O*_{A1} *carro*_{A1} *que os*_{A2} *mecânicos*_{A2} *consertaram*_{B2} *bateu*_{B1}; e DNAs do tipo *alinhadas* (A1 A2 B1 B2), em que os elementos dependentes estão em sequência imediata, como em *O*_{A1} *mecânico*_{A1} *consertou*_{A2} *os*_{B1} *carros*_{B1} *que bateram*_{B2}. Resultados de dois experimentos com a tarefa de labirinto (*maze task*) revelaram tempo de leitura maior na condição aninhada, sugerindo que dependências nessa configuração causam maior demanda no processamento linguístico. As DNAs foram bastante exploradas no âmbito da aquisição da linguagem, mas pouco se sabe a respeito da habilidade de processamento dessas estruturas em adultos. O estudo em questão lança luz a respeito da discussão do processamento de múltiplas DNAs por adultos no PB e sobre os fatores envolvidos no processamento de múltiplas dependências.

Palavras-chave: Dependências não adjacentes; múltiplas dependências; processamento linguístico; concordância.

ABSTRACT

In this paper, we investigate the processing of nonadjacent dependencies (NADs) and their different configurations in Brazilian Portuguese (BP) by adult speakers. NADs are defined as the co-occurrence of morphosyntactic and/or phonological markers with intervening material, as in *A* *menin-a*, in which the dependence between the determiner and the *a-* noun gender feminine mark is observed, with the intervening material *menin-* between them. From a structural point of view, there are nested NADs (A1 A2 B2 B1), when the co-dependent elements A2 and B2 are “nested” between A1 and B1, as in *The*_{A1} *car*_{A1} *the*_{A2} *mechanics*_{A2} *repaired*_{B2} *crashed*_{B1}; and aligned NADs (A1 A2 B1 B2) when the co-dependent elements are in immediate sequence, as in *The*_{A1} *mechanic*_{A1} *repaired*_{A2} *the*_{B1} *cars*_{B1} *that crashed*_{B2}. Results from two experiments using maze tasks revealed longer reading times for the nested condition, suggesting that dependencies in this configuration cause greater demand in linguistic processing. NADs have been explored in language acquisition, but little is known about how well adults process these structures. This study sheds light on adult processing of multiple NADs in BP and on the factors involved in multiple dependency processing.

Keywords: Nonadjacent dependencies; multiple dependencies; linguistic processing; agreement.

1 O projeto de pesquisa ao qual o presente artigo se vincula teve apoio da FAPEMIG (Processo nº. APQ-00988/15) e conta com a aprovação do Comitê de Ética em Pesquisa/UFJF (CAAE: 44123015.6.0000.5147).

2 Doutoranda em Linguística pelo Programa de Pós-graduação em Linguística da Universidade Federal de Juiz de Fora (UFJF). E-mail: kessiasilvahenrique@gmail.com.

3 Doutora em Letras pela PUC-Rio. Professora da Universidade Federal de Juiz de Fora (UFJF). Bolsista de Produtividade do CNPq. E-mail: cristina.name@ufjf.edu.br.

1. Introdução

O presente trabalho busca investigar o processamento de múltiplas dependências não adjacentes a partir de suas possíveis configurações e considera como elas podem influenciar o processamento de sentenças por adultos falantes nativos do português brasileiro (doravante, PB).

Dependências não adjacentes (DNAs) são caracterizadas como padrões de co dependência de elementos – itens funcionais, em sua maioria – que ocorrem em posição não adjacente, i.e., com material interveniente entre eles (NAME; LAGUARDIA, 2013; SANDOVAL; GÓMEZ, 2013; LAGUARDIA, 2016). Tais dependências expressam relações estruturais, de concordância entre um elemento controlador e um elemento por ele controlado, que tem sua forma morfofonológica determinada pelo estabelecimento dessa operação, como ilustram os seguintes exemplos:

(1) **Eu**_(pronome) estud-**o**_(terminação verbal)

(2) **O**_(determinante) cadern-**o**_(vogal temática)

No exemplo (1), acima, observa-se uma relação de co dependência entre o pronome *eu* e a terminação verbal *-o*, que marca a desinência de primeira pessoa do singular, sendo que tais elementos encontram-se linearmente separados pelo material interveniente que é a base verbal *estud-*. No exemplo (2) também há dependência entre elementos, dessa vez entre o determinante *o* e a vogal temática *-o* do nome, cuja base *cadern-* intervém entre os elementos co dependentes.

Os dois exemplos ilustram relações de concordância no PB, sendo o primeiro de concordância verbal (sujeito-verbo) e o segundo, de concordância nominal de gênero. As DNAs, portanto, podem ser entendidas como manifestações de regras abstratas de concordância. Essa distinção entre concordância – em uma perspectiva gerativista de língua, operação sintática de checagem (CHOMSKY, 1995) ou de valoração de traços (CHOMSKY, 1999, 2001) – e DNA se faz importante, tanto nos estudos em aquisição da linguagem quanto em processamento adulto, na medida em que a mesma regra de concordância – verbal, por exemplo, pode ser expressa por diferentes elementos em covariação (diferentes formas pronominais e afixos verbais, p.ex.) e sob diversas configurações (p.ex., com maior ou menor distância entre os elementos codependentes). Segundo De Vries (2011, p.2065), o processamento de DNAs pode ser considerado um marco para a linguagem humana, uma vez que permite que padrões sejam reconhecidos, disparando mecanismos de concordância.

No processo de aquisição de uma língua, o reconhecimento dos padrões de DNAs permitiriam ao aprendiz identificar as regras de concordância subjacentes à língua. De fato, o mapeamento de DNAs

e seus possíveis padrões em uma dada língua tem recebido destaque nos estudos que investigam os mecanismos subjacentes à aquisição (ou aprendizagem) de língua, seja por bebês (GÓMEZ; MAYE, 2005; NAME; SHI; KOULAGUINA, 2011; LAGUARDIA; NAME; SHI, 2013; LAGUARDIA et al., 2015, entre outros) ou por adultos (DE VRIES et al., 2012; VAN DEN BOS; CHRISTIANSEN; MISYAK, 2012; ZUHURUDEEN; HUANG, 2016; entre outros).

No que se refere ao processamento adulto de L1, uma questão relevante é se haveria diferenças de processamento em função da configuração em que elas se manifestam. As DNAs podem assumir diversas configurações e não é rara a ocorrência de múltiplas DNAs em uma sentença e, ainda, de DNAs embutidas umas nas outras. Do ponto de vista estrutural, essas relações podem ser de vários tipos e conter múltiplas dependências. Por exemplo, há DNAs do tipo *aninhadas* (*nested*: A1 A2 B2 B1), em que os elementos codependentes A2 B2 “se aninham” entre A1 B1; do tipo *alinhadas* (ou próximas: A1 A2 B1 B2), em que os elementos em concordância estão em sequência imediata; e DNAs cruzadas, em que as dependências se cruzam (*crossed*: A1 B1 A2 B2), conforme os exemplos apresentados abaixo em inglês (3), holandês (4) e PB (5):⁴

(3) [**All the animals**]_{A1} (that) [**Helena**]_{A2} *loves*_{B2} **are**_{B1} wild.⁵
[Todos os animais] (que) Helena adora são selvagens.
Todos os animais que Helena adora são selvagens.

(4) (dat) [**Jan**]_{A1} [**Marie**]_{B1} **zag**_{A2} *zwemmen*_{B2}.⁶
(que) Jan Marie viu nadar.
(que) Jan viu Marie nadar.

(5) [**O pedreiro**]_{A1} **construiu**_{A2} *o prédio*_{B1} que *desabou*_{B2}.⁷

Em (3), há um exemplo de DNAs aninhadas, em que a DNA *Helena loves* está aninhada entre a DNA *all the animals ... are*. No exemplo (4) temos a ocorrência de DNAs em uma configuração cruzada, de forma que as DNAs *Jan...zag* perpassam pelas DNAs *Marie...zwemmen*. Em (5), as DNAs estão em uma sequência imediata, em uma configuração alinhada.

4 Vale ressaltar que nos exemplos estamos destacando o morfema verbal e o DP como um todo, pois o foco do trabalho são as configurações das dependências. No entanto, essas relações de dependência também ocorrem no interior do DP.

5 Fonte: elaboração própria.

6 Fonte: M.H. de Vries et.al., 2011, p. 2066.

7 Fonte: elaboração própria.

Ainda pouco se sabe sobre o processamento adulto diante das diversas possibilidades de DNAs que as línguas naturais apresentam. Nesse sentido, torna-se relevante que esse fenômeno seja explorado em todas suas as dimensões nas mais diversas línguas.

Neste artigo, focalizaremos as relações estruturais expressas pelas DNAs. No PB, é mais comum encontrar a ocorrência de múltiplas DNAs em orações relativas. Dessa forma, abordaremos as múltiplas dependências a partir das orações relativas, em sentenças como *O carro que os mecânicos consertaram bateu*. Na seção a seguir exemplificaremos como essas configurações podem se manifestar em uma língua natural.

Primeiramente, apresentaremos as possíveis configurações que as DNAs podem assumir nas mais diversas línguas, inclusive no PB. Já na seção 3, faremos uma breve revisão sobre o processamento de orações relativas no PB e teorias que exploram a dificuldade no processamento de alguns tipos de relativas, uma vez que em algumas configurações de orações relativas no PB há múltiplas dependências e essas dependências se encontram em uma configuração aninhada. Na seção 4, são reportados dois experimentos desenvolvidos com o objetivo de investigar o processamento adulto de múltiplas DNAs no PB. Finalmente, na seção 5, serão apresentadas as considerações finais feitas a partir da investigação proposta neste estudo.

2. As possíveis configurações de DNA

DNAs podem ser expressas de diversas maneiras nas mais diferentes línguas e podem assumir diversas configurações. Uma das configurações possíveis no PB contendo múltiplas dependências é quando há a sequência imediata dos elementos $A_1B_1A_2B_2$ como em:

(6) **O mecânico consertou os carros que bateram.**



Além de estar em sequência imediata, as DNAs podem assumir outras configurações. Na configuração chamada de dependências aninhadas (ou *nested*, no inglês), as dependências estão umas “dentro” das outras, como no esquema $A_1A_2B_2B_1$, em que o elemento A_1 é dependente do elemento B_1 , exemplificado a seguir:

(7) **Os materiais que o diretor encomendou chegaram.**



Em línguas como holandês e alemão, como exemplifica De Vries *et al* (2011), há a possibilidade

de dependências cruzadas (ou *crossed*, no inglês). Nesse caso, as dependências entre elementos vão cruzando umas às outras, como no esquema $A_1A_2B_1B_2$, em que A_1 é dependente de B_1 e A_2 de B_2 , como exemplifica o exemplo em holandês:

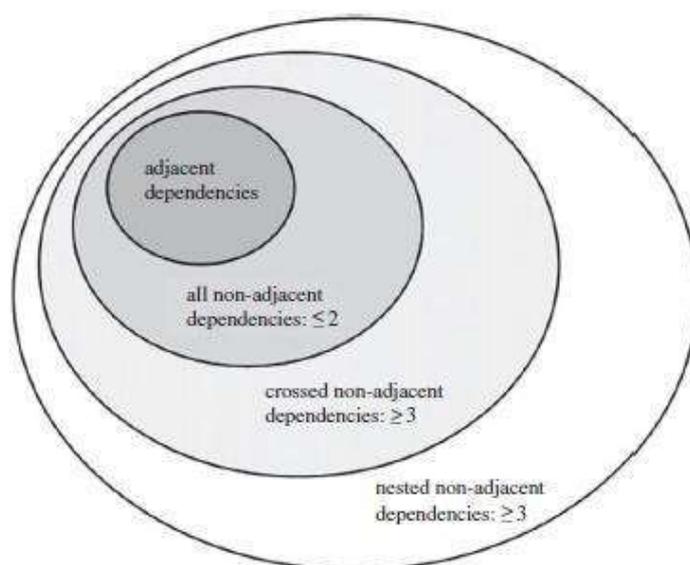
(8) Jan Marie zag zwemmen.



(que) Jan Marie viu nadar.

Estudos realizados a partir de uma gramática artificial (De Vries *et. al* 2012, 2011) sugerem que a existência de múltiplas dependências é um fator que aumenta o custo do processamento. De Vries *et al* (2012) propõem uma hierarquia de complexidade no processamento de múltiplas dependências:

Figura 1 – Hierarquia de complexidade no processamento de múltiplas dependências porposta por De Vries *et. al.* (2012).



Fonte: De Vries *et. al.* (2012, p. 2068)

De acordo com o gráfico proposto, dependências adjacentes seriam as dependências que causariam menos demanda cognitiva no processamento. Nesse sentido, dependências em relações de adjacência seriam as menos custosas ao processamento em comparação a todas as dependências não adjacentes. Essa visão é corroborada pelos resultados do estudo de Gómez (2002; GÓMEZ; MAYE, 2005), com bebês entre 12 e 18 meses expostos a estímulos de uma língua artificial. A autora defende que, durante o processo de aquisição, dependências adjacentes seriam adquiridas mais precocemente se comparadas com as dependências que não estão em relação de adjacência. É importante observar, no entanto, que dependências que sejam de fato adjacentes, ou seja, cujos itens dependentes estejam em sequência imediata, parecem ser possíveis apenas em estímulos criados a partir de línguas artificiais.

O que muitas vezes é chamado de dependências adjacentes em línguas naturais são, na verdade, dependências não adjacentes, uma vez que há material interveniente entre os termos dependentes – a base ou raiz do item lexical (p.ex., *o carr-o, eu danç-o*). Podemos encontrar relações de adjacência apenas em línguas artificiais. Gómez (2002), por exemplo, investigou a variabilidade e detecção de estruturas invariáveis, criando estímulos que estavam em relação de adjacência, de forma que todas as vezes que o participante ouvia, por exemplo, *PEL*, imediatamente após ouvia *WADIM*, de forma que esses itens estavam em relações de dependência imediata, sem nenhum material interveniente.

Além disso, ainda segundo De Vries *et al.* (2011), as DNAs não seriam processadas da mesma forma. Dependências em uma configuração cruzada com mais de três elementos dependentes causariam menos demanda cognitiva do que dependências em uma configuração aninhada com mais de três elementos dependentes. Assim, o processamento de DNAs do tipo cruzadas, como no exemplo (8) demandariam menos custo que o de dependências aninhadas e a quantidade de dependências na sentença a serem resolvidas simultaneamente não poderia exceder a três. Nesse sentido, como apontado por De Vries *et al.* (2012), dois fatores são relevantes no que diz respeito à complexidade do processamento. O primeiro deles é a ordem ou a configuração em que as DNAs estão dispostas em uma sentença e o segundo é a quantidade de dependências que precisam ser resolvidas quase que simultaneamente.

No PB, a ocorrência de múltiplas DNAs é mais recorrente em orações relativas. Ainda seguindo a escala proposta por De Vries *et al.* (2012), podemos encontrar DNAs em uma configuração aninhada, como no exemplo (7) e DNAs alinhadas, como no exemplo (6). No que tange às DNAs do tipo cruzadas, não há ocorrências dessas dependências no PB, apenas? em línguas como o alemão e o holandês (DE VRIES *et al.*, 2012), conforme em (8).

A próxima seção revisitará alguns estudos a respeito do processamento de orações relativas e sobre quais configurações são reportadas na literatura como geradoras de maior dificuldade no processamento, assim como os principais fatores sugeridos como causadores da dificuldade. Usamos como estrutura para estudar o fenômeno do processamento das múltiplas DNAs as orações relativas pelo fato de que é nas orações relativas que o fenômeno de múltiplas dependências fica mais evidenciado. Talvez, em alguns casos, a dificuldade do processamento de algumas orações relativas se dê pelo fato de haver múltiplas dependências de diferentes configurações na sua estrutura.

3. O processamento de orações relativas – breve revisão

A pesquisa sobre orações relativas em uma perspectiva psicolinguística tem investigado, ao longo dos anos, as diferentes configurações que essas orações podem assumir e as implicações dessas

diferentes configurações na aquisição da linguagem e no processamento adulto. De acordo com Miranda (2008), questões referentes à compreensão e à aquisição dessas estruturas, considerando foco e encaixamento como problemas relativos ao custo do processamento, têm sido amplamente investigadas nos últimos anos. Ainda de acordo com Miranda (2008, p.45), o problema do foco estaria ligado “à função do vestígio interno (cópia) na relativa, caracterizada como *gap* em modelos psicolinguísticos”. O problema do encaixamento, por sua vez, diz respeito à função sintática do DP em que se insere a oração relativa. Nessa perspectiva, a oração relativa pode ser encaixada no DP que é sujeito (exemplos **a** e **b** abaixo) ou pode ser encaixada à direita da oração principal nos DPs que são complementos, como objeto direto (exemplos **c** e **d** abaixo), por exemplo.

A partir da manipulação de foco e encaixamento, Miranda (2008) aponta que são encontrados diferentes tipos de relativas, mas a literatura psicolinguística tem focado em quatro tipos principais, sendo eles (MIRANDA, 2008, p.45):

- a- Relativas SS que modificam o sujeito da oração principal e o *gap* na relativa tem função de sujeito – A estudante [que irritou a professora] saiu da escola.
- b- Relativas SO que são encaixadas na posição de sujeito da oração principal e *gap* com função de objeto – A estudante [que a professora irritou] saiu da escola.
- c- Relativas OS que são ramificadas em posição de objeto na oração principal e tem foco no sujeito – O diretor convocou a aluna [que desrespeitou o professor].
- d- Relativas OO em que o elemento modificado na oração principal é um objeto e a função do *gap* também é de objeto – O diretor convocou a aluna [que o professor reprovou.]⁸

De uma maneira geral, no que diz respeito à compreensão de orações relativas, a literatura aponta para um maior custo no processamento de relativas SO, ou seja, aquelas encaixadas na posição de sujeito da oração principal. Várias propostas buscaram explicar por que certas estruturas, que contêm a mesma quantidade de informação que outras, podem gerar maior dificuldade. Muitas dessas propostas são fundamentadas na limitação da memória de trabalho e aspectos estruturais e/ou sintáticos.

De maneira breve, abordaremos algumas propostas que buscam explicar a maior dificuldade na compreensão de relativas do tipo SO. A primeira proposta é a que foi chamada de *Canonical Word Order*, de Bever (1970), segundo a qual existiriam interpretações preferíveis a outras e o *parser* estabeleceria estratégias que dariam suporte a expectativas sintáticas e semânticas no decorrer do

8 Exemplos retirados de MIRANDA (2008, p. 46)

processamento e, de acordo com essas expectativas, seria construída uma estrutura básica para atribuir papel temático e função sintática aos elementos da sentença. A sequência N (nome) V (verbo) N (nome) é identificada como oração principal e a essa ordem sintática é atribuído o papel temático agente-ação-paciente, o que favorece o processamento de relativas SS, que modificam o sujeito da oração principal e o *gap* na oração relativa tem a função de sujeito.

A teoria chamada *Late Assignment of Syntax Theory*, apresentada por Townsend e Bever (2001), propõe um modelo de compreensão baseado em dois estágios. No primeiro estágio, intitulado como pseudossintaxe, o *parser* usaria a ordem canônica da língua, que no PB seria SVO. Essa ordem é estabelecida como um *template* para representar relações gramaticais e relações de papéis temáticos que seriam *default* na língua. O segundo estágio, denominado Sintaxe Real, faria reanálises para construir estruturas frasais. Nesse sentido, relativas OS demandariam maior custo no processamento porque não seguem a mesma sequência da ordem canônica da língua, pois tem a ordem OSV.

Já na teoria *Parallel Function*, de Sheldon (1974), baseada em fatores sintáticos, a discrepância no processamento entre relativas com foco no sujeito e foco no objeto seria atribuída a diferentes papéis que o núcleo nominal desempenharia na oração principal em contraste com seu papel na oração relativa. Nesse sentido, em orações relativas do tipo SS, o *gap* possui a mesma função sintática que o *filler*, de forma que os dois são sujeitos da sentença. Nas orações relativas SO, em que a relativa é encaixada no sujeito e o foco é no objeto, o *filler* tem função sintática de sujeito, ao passo que o *gap* tem função sintática de objeto. Assim, o mesmo núcleo nominal receberia dois papéis temáticos diferentes. No entanto, McWhinney e Pleh (1988) ampliaram essa teoria sugerindo a *Perspective Shift*, em que o *default* para o *parser* seria assumir o ponto de vista de agente. Segundo Miranda (2008, p.48) “considera-se que o processamento é mais fácil quando a mesma perspectiva é mantida”, o que não aconteceria nas orações SO, em que a troca de perspectiva é necessária.

A teoria proposta por Pritchett (1992), *Grammatical Theory of Processing*, prevê que o *parser* atribui papel temático ao núcleo nominal da oração o quanto antes, entretanto, nas relativas SS, quando o verbo da oração relativa é analisado, o *parser* consegue atribuir papel temático para o núcleo nominal na posição de sujeito, mas o núcleo nominal que ocupa a posição de objeto só recebe papel temático depois que o verbo já atribuiu o papel temático para o sujeito da oração relativa. Isso também explicaria a dificuldade em orações com foco no objeto (OS) em comparação com aquelas com foco no sujeito (SS).

Além de todas essas propostas baseadas na estrutura sintática das orações, também existem modelos que explicam a maior dificuldade em orações OS com base em restrições na memória de trabalho (FRAZIER; FODOR 1978; MACWHINNEY, 1987) uma vez que o *filler* precisa ser mantido por mais tempo na memória até encontrar com o *gap* e só assim o *gap* receberá caso e papel temático diferentes do *filler*. Em resumo, relativas OS, com encaixamento no sujeito e foco no objeto seriam

mais custosas em termos de processamento.

No que diz respeito à compreensão dessas estruturas, alguns fatores foram explorados e são sugeridos como causadores da maior dificuldade no processamento de relativas de objeto. Um dos fatores sugeridos é a distância existente entre o *filler* e o *gap*. Vale ressaltar que a interferência do fator distância já foi amplamente explorado na literatura sociolinguística no que diz respeito à concordância e também, no processamento adulto de sentenças no PB, em um trabalho recente de Henrique (2013). A autora investigou a distância linear no processamento da concordância verbal contrastando sentenças como *Os vestidos **custam** barato no shopping* e *Os vestidos na liquidação **geralmente custam** barato no shopping*. Os resultados sugerem que esse fator exerce influência nessa operação sintática, uma vez que quanto mais material interveniente entre sujeito e verbo, maior o custo do processamento. Alguns autores exploraram esse fator no processamento de orações relativas de objeto e os resultados sugerem que a distância também é um fator importante no que diz respeito ao processamento de orações relativas (FIEBACH; SCHELESEWSKY; FRIEDERICI, 2002). Uma explicação para tais resultados seria que o *gap* deve ser mantido mais tempo na memória até encontrar o elemento que o subcategoriza, resultando em uma maior demanda nos custos do processamento da sentença.

Outras teorias que levam em consideração dificuldades impostas pela categoria dos núcleos nominais também foram propostas, como a teoria chamada de *similarity-based interference* proposta por Gordon, Hendrick e Johnson (2001). De acordo com essa teoria, núcleos nominais de categorias diferentes tornariam o processamento menos custoso, ao passo que núcleos nominais pertencentes a uma mesma categoria dificultaria o processamento. Isso aconteceria porque quando os núcleos são da mesma natureza, haveria uma certa “confusão” decorrente da semelhança, dificultando a recuperação do núcleo nominal, no entanto isso aconteceria em relativas de objeto, mas não nas relativas de sujeito.

Além de fatores sintáticos, fatores de ordem semântica, como a animacidade dos núcleos nominais da oração principal e da oração relativa, têm sido considerados. De acordo com experimento realizado por Traxler *et al.* (2002) usando *eye-tracking*, quando dois referentes podem ser agente do verbo, a dificuldade no processamento aumenta, ou seja, quando apenas um referente é plausível de ser o agente do verbo, o custo de processamento é menor. No entanto, no que diz respeito à semântica, não é apenas a possível atribuição de papel temático que pode dificultar ou facilitar o processamento. De acordo com Traxler *et al.* (2002), a animacidade dos núcleos nominais também parece ser um fator relevante. Em experimento realizado no estudo desenvolvido por Traxler *et al.* (2002), quando o sujeito

da oração principal era animado e o sujeito da relativa inanimado em relativas com encaixamento no sujeito e foco no objeto, o processamento foi mais lento, mas quando o oposto acontecia, ou seja, sujeito da oração principal inanimado e sujeito da encaixada animado, o processamento era facilitado.

Nessa breve revisão, vimos que a ordem linear e a distância entre os elementos correlacionados são fatores considerados em modelos que buscam dar conta de diferenças no processamento de relativas de variadas configurações. De acordo com a literatura, a animacidade também parece ser um fator relevante para o processamento da concordância. Na seção a seguir apresentaremos duas atividades experimentais desenvolvidas a fim de explorar empiricamente o processamento de múltiplas DNAs em diversas configurações por meio de orações relativas no PB.

4. O processamento de múltiplas DNAs no PB: dados experimentais

O objetivo dos experimentos que serão descritos nesta seção foi investigar em que medida a configuração das DNAs e a manipulação da informação de número do sujeito se mostram relevantes no processamento da concordância no PB, uma vez que as DNAs também expressam relações estruturais de concordância de número.⁹

4.1 Método

Nas atividades experimentais que serão descritas a seguir foi utilizada uma variante do paradigma de leitura monitorada, implementado por meio de uma tarefa de labirinto (*maze task*), utilizando-se o *software Linger*. A escolha dessa técnica experimental se justifica pelo fato de esse tipo de tarefa oferecer uma alternativa às tarefas clássicas de leitura automonitorada. A tarefa de labirinto apresenta uma vantagem em comparação a outros procedimentos de leitura, pois força a integração incremental de cada novo segmento lido com o contexto prévio. Assim, essa técnica pode fornecer medidas mais precisas do processamento dos segmentos analisados.

Tarefas de labirinto permitem medir o tempo de leitura de cada segmento, além de possibilitar a identificação das escolhas dos participantes em cada ponto do estímulo. A grosso modo, funcionam de forma que o participante vê duas palavras ou expressões por vez separadas por barras. Apenas uma delas é a opção correta que dá sequência à sentença de forma coerente. O participante deve selecionar

⁹ As atividades experimentais descritas foram aplicadas a estudantes universitários, com alto grau de instrução formal. Além disso, o experimento foi conduzido em uma instituição de ensino, o que gera aumento no monitoramento do nível da linguagem. A existência da possibilidade da concordância variável não interferiu nos resultados, uma vez que se essa questão tivesse sido relevante no design experimental, ela refletiria em todas as condições experimentais, o que não ocorreu. Agradecemos aos revisores por chamarem atenção para essa possibilidade, visto que a concordância variável é uma possibilidade no PB.

através de botões se a palavra ou o segmento que julga ser mais adequado é o que está à direita da tela ou à esquerda. O participante chega ao final da sentença selecionando os itens que compõem a frase em cada passo.

Design Experimental: as variáveis independentes foram configuração da dependência não adjacente (alinhada vs. aninhada) e número do sujeito (singular vs. plural), obtendo-se, assim, o design 2x2. As variáveis dependentes foram tempo de escolha/leitura de cada segmento e número de escolhas dos segmentos alvo em cada condição.

Na atividade experimental descrita, os participantes foram expostos a um conjunto de 20 sentenças, contendo 12 sentenças experimentais, sendo metade delas com sujeito na 3ª pessoa no singular e metade com sujeito na 3ª pessoa no plural, além de 20 distratoras de diversos tamanhos e configurações e 2 sentenças para fase de treinamento e habituação com a tarefa. Vale ressaltar que utilizamos um número menor de sentenças distratoras do que o habitual porque nossas condições experimentais apresentam uma variação considerável entre si no que diz respeito a configuração e tamanho.

As sentenças experimentais foram distribuídas em quadrado latino que era gerado pelo próprio *software* utilizado na programação e aplicação do teste. As sentenças experimentais foram classificadas entre os estímulos conforme a configuração das DNAs em cada sentença, a saber: aninhada – quando havia uma DNA embutida em outra DNA, e alinhada – quando uma DNA aparecia seguida de outra DNA.

A seguir, apresentamos exemplos dos estímulos experimentais em cada uma das condições investigadas.

Quadro 1 – Condições experimentais.

NÚMERO DO SUJEITO	CONFIGURAÇÃO DA DNA	
	ALINHADA	ANINHADA
SINGULAR	O engenheiro construiu <i>os prédios</i> que <i>desabaram</i> na cidade.	O prédio que <i>os engenheiros construíram</i> desabou na cidade.
PLURAL	Os engenheiros construíram <i>o prédio</i> que <i>desabou</i> na cidade.	Os prédios que <i>o engenheiro construiu</i> desabaram ontem.

Fonte: elaboração própria

Com a intenção de não adicionar mais uma variável ao design experimental, controlamos a saliência fônica usando apenas verbos no pretérito, uma vez que esse tempo verbal apresenta menor

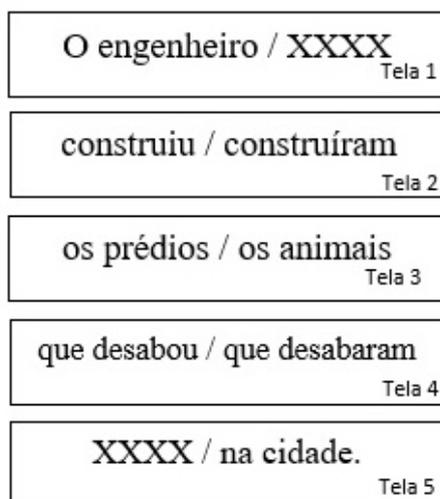
grau de saliência fônica. Também controlamos o número de sílabas dos sujeitos e dos verbos, de forma que ficasse equilibrado em todas as sentenças experimentais. Os estímulos foram organizados em quatro listas distribuídos em um quadrado latino gerado pelo próprio software utilizado, de forma que cada participante leu uma versão dentre as quatro possibilidades de uma mesma sentença e nenhuma sentença foi repetida em mais de uma condição para o mesmo participante. Os participantes foram aleatoriamente distribuídos nas quatro listas.

Previsões: (I) A condição DNA aninhada deve registrar o maior tempo de escolhas dos verbos alvo, principalmente do segundo verbo, em comparação com a condição alinhada. Essa previsão decorre do fato de uma possível sobrecarga na memória de trabalho no processamento de sentenças nessa configuração, dificultando a recuperação do número do sujeito e, conseqüentemente, aumentando o custo do processamento da sentença, uma vez que o participante gastaria um tempo maior na tomada de decisão, pois a distância estrutural entre os itens dependentes é maior; (II) no que diz respeito à escolha do segmento alvo, espera-se que o segundo verbo da condição aninhada registre taxas de escolha do verbo-alvo menores, uma vez que a complexidade do material interveniente possa favorecer uma eventual perda da informação de número gramatical do sujeito da oração principal e, como resultado, o participante aceitaria o verbo com número incongruente do sujeito. Em contraponto, a condição alinhada apresentaria maior número de acertos do segmento alvo, uma vez que os itens dependentes estão em sequência imediata.

Participantes: O experimento foi integralmente conduzido na Faculdade de Letras na Universidade Federal de Juiz de Fora, totalizando 24 participantes voluntários, com idade entre 17 e 37 anos (idade média: 19,2 anos). Nenhum participante ou dado foi excluído.

Procedimento: Os participantes foram testados individualmente, sendo instruídos a se sentarem em frente ao *notebook*. As instruções para realização da tarefa eram apresentadas por escrito na tela do computador e também explicadas oralmente pelo pesquisador antes do início da tarefa. As instruções informavam que o participante seguiria uma sentença através de um labirinto, seria apresentado a duas palavras ou segmentos por vez na tela; esses segmentos eram separados por barras e também era informado que só uma das opções daria sequência à sentença de forma coerente. Em cada passo, o participante deveria selecionar uma das palavras/segmentos apresentados utilizando um dos dois botões que estavam disponíveis e sinalizados no teclado. Esses botões eram a letra C e M no teclado, marcados respectivamente com ← (opção da esquerda) e → (opção da direita). A figura abaixo ilustra o procedimento, sendo que cada par de palavras ou segmentos aparecia para os participantes em uma tela independente.

Figura 2 – Exemplo do procedimento da tarefa de labirinto.



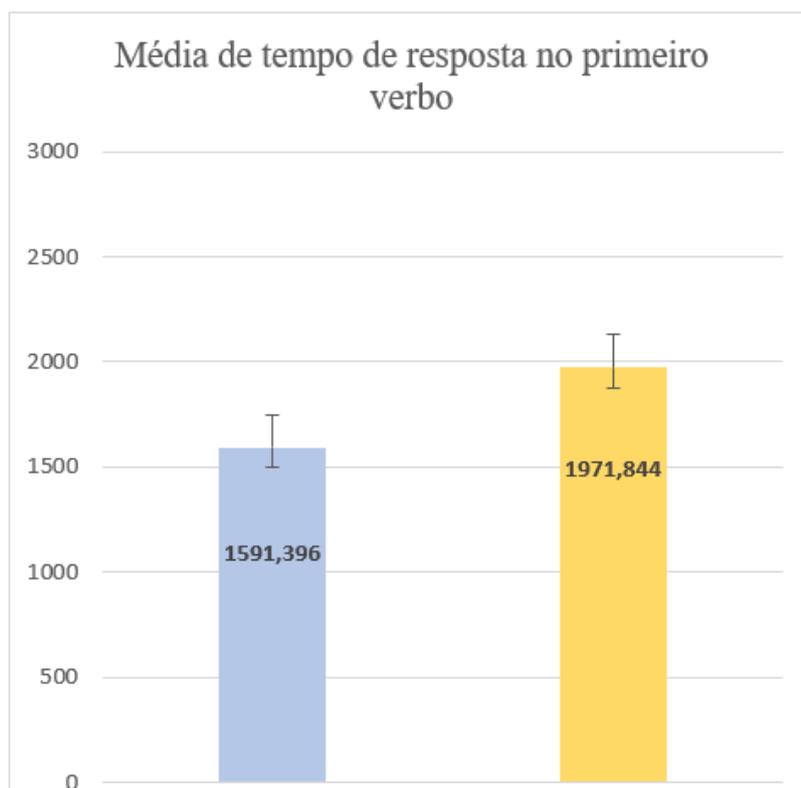
Após ler e ouvir as instruções, os participantes foram submetidos a uma pequena fase de treinamento contendo duas sentenças. O objetivo do treinamento era fazer com que os participantes se familiarizassem com a tarefa experimental. Após o começo da atividade, o participante controlava a velocidade de aparição de novos estímulos na tela conforme escolhia entre duas opções que eram exibidas até completar uma frase (no final da frase havia um ponto final, de forma que o participante sabia que aquele era o final e que em seguida uma nova frase começaria). A duração de cada sessão experimental foi de aproximadamente 6 minutos.

4.2 Resultados e discussão

Para a variável *tempo de resposta*, foi considerado o tempo registrado para as escolhas em que havia verbo em cada condição. Os resultados foram submetidos ao teste de igualdade de variâncias do erro de Levene no *software Statistical Package for the Social Sciences* (SPSS) e os resultados revelaram homogeneidade da amostra, tanto no que se refere aos dados obtidos no primeiro verbo quanto no segundo verbo. Dessa forma, os dados foram submetidos a uma análise da variância (ANOVA) de design fatorial 2x2 de medidas repetidas (*within-subjects*). No que diz respeito ao tempo de resposta do primeiro verbo, foi obtido um efeito principal na variável *configuração de DNA* ($F(1,47) 8,40 = p < 0,005$), com médias maiores nas condições com configuração aninhada. Não foram encontrados efeito principal da variável *número do sujeito* nem efeito de interação. Foram realizados testes *post hoc* para comparações entre pares das condições (teste-t pareado, unicaudal), que confirmaram a diferença estatisticamente significativa entre as condições DNA alinhada ($M=1591,39ms$, $DevP=641,85ms$) vs.

DNA aninhada ($M=1971,84\text{ms}$, $\text{DevP}=1099,77\text{ms}$; $t(23)=2,64$, $p=0,007$). Além disso, nenhum dado foi excluído nos resultados. O gráfico abaixo ilustra as médias de tempo de resposta nas condições alinhada e aninhada:

Gráfico 1: média do tempo de reação em cada condição (primeiro verbo)



Fonte: elaboração própria

No quadro abaixo especificamos as médias de tempo de respostas no primeiro verbo em cada condição experimental:

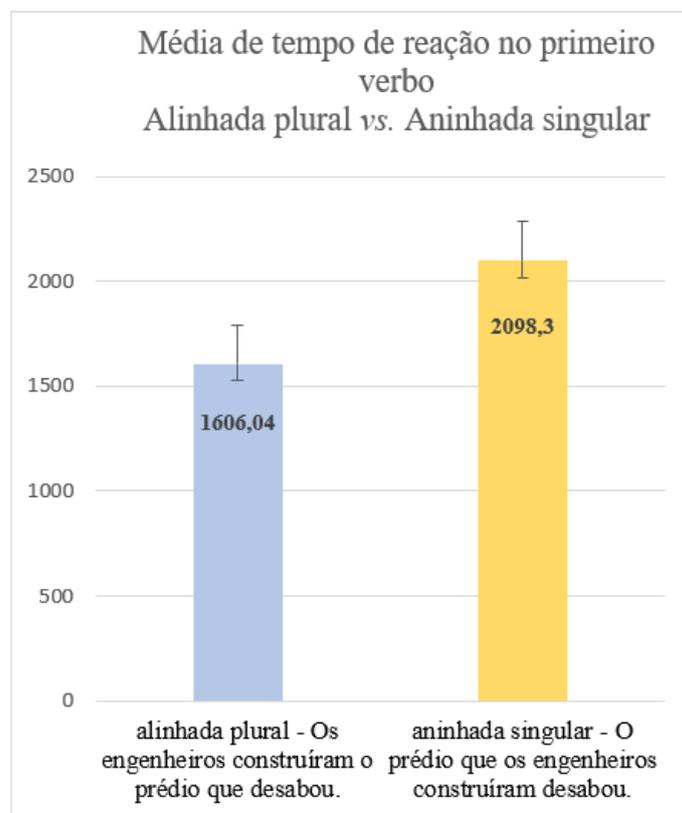
Quadro 2: média do tempo de reação no primeiro verbo em todas as condições experimentais

NÚMERO DO SUJEITO	Médias de tempo de resposta em cada condição – primeiro verbo	
	ALINHADA	ANINHADA
SINGULAR	1576,31ms	2098,38ms
PLURAL	1606,48ms	1845,31ms

Fonte: elaboração própria

Houve também diferença significativa no par ‘alinhada plural’ ($M=1606,48\text{ms}$, $\text{DevP}=559,96\text{ms}$) vs. ‘aninhada singular’ ($M=2098,38\text{ms}$, $\text{DevP}=1293,87\text{ms}$; $t(23)=2,13$, $p=0,03$).

Gráfico 2: média do tempo de reação nas condições alinhada plural e aninhada singular (primeiro verbo)



Fonte: elaboração própria

No que diz respeito ao segundo verbo, não foram verificados efeitos principais ou de interação dos fatores; também não houve diferença significativa nas comparações entre as condições. O quadro abaixo reporta as médias de tempo de reação obtidas no segundo verbo.

Quadro 3: média do tempo de reação no segundo verbo em todas as condições experimentais

NÚMERO DO SUJEITO	Médias de tempo de resposta em cada condição – segundo verbo	
	ALINHADA	ANINHADA
SINGULAR	2089,54ms	1969,79ms
PLURAL	1840,71ms	2125,81ms

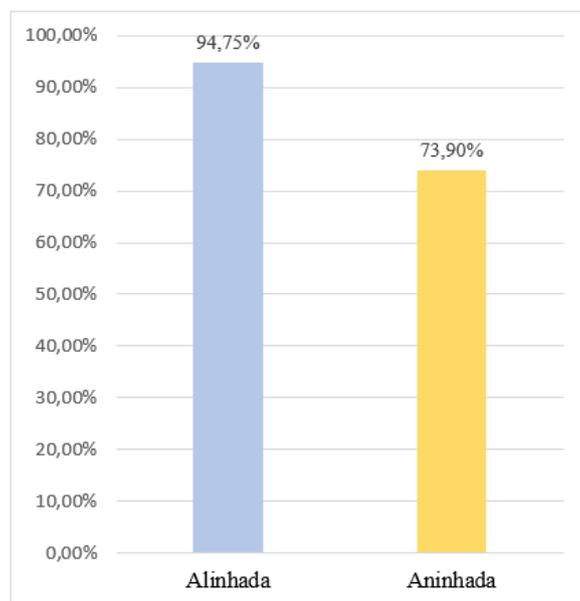
Fonte: elaboração própria

No que tange à segunda variável, número de acertos, os dados foram submetidos ao teste não paramétrico *Q de Cochran*, que faz o tratamento de dados dicotômicos e, posteriormente, foi realizado o teste *McNemar*, que além de fazer análise de dados dicotômicos também realiza a comparação entre pares. Analisando as taxas de acerto no primeiro verbo, não houve efeito estatisticamente relevante. As taxas de acerto foram 88% na condição alinhada plural, 100% na condição alinhada singu-

lar, 87,5% na condição aninhada plural e 87,5% na condição aninhada singular.

Entretanto, o segundo verbo registrou taxas de acertos baixas nas condições aninhadas singular (79,1%) e plural (68,7%) em comparação com as condições alinhadas singular (95,8%) e plural (93,7%) (Q de Cochran 21,959^a=p<,0000):

Gráfico 3: Taxa de acertos do segmento alvo (segundo verbo)



Fonte: Elaboração própria

Os resultados obtidos no experimento são compatíveis com a ideia de que o fator configuração da DNA seria relevante para o custo do processamento e realização da concordância e que as duas configurações de DNA possíveis no PB (aninhada – alinhada) não seriam igualmente processadas pelos falantes. A condição aninhada se mostrou um contexto mais custoso para o processamento, tal como evidenciam os tempos de reação maiores e as taxas de acerto menores registradas nessa condição, uma vez que estruturalmente essas sentenças são mais complexas. No primeiro verbo, houve efeito principal de configuração de DNA e diferença estatisticamente significativa entre as condições *alinhada plural* e *aninhada singular*. Ambos os verbos estavam no plural e próximos de seu sujeito (*Os engenheiros construíram o prédio que desabou.* vs. *O prédio que *os engenheiros construíram desabou.*), o que pode sugerir que a presença de uma dependência embutida pode causar interferência no mapeamento entre sujeito e verbo independente da informação de número ou da distância linear entre eles. Esses resultados são compatíveis com nossa previsão de maior tempo de escolha do verbo alvo para a condição DNA aninhada, mas esperávamos que isso acontecesse particularmente para o segundo verbo, o que não aconteceu. Os tempos médios de escolha do primeiro verbo foram menores na condição DNA alinhada do que na DNA aninhada, tanto no singular (1576,31ms vs. 2098,38ms) quanto no plural (1606,48ms vs. 1845,31ms). Já para o segundo verbo,*

as médias de tempo de reação foram todas muito próximas entre si e a condição DNA aninhada singular teve o segundo menor tempo médio entre as condições (1969,79ms), ficando acima apenas da condição DNA alinhada plural (1840,71ms). Esses resultados são difíceis de serem explicados e, até o momento, não encontramos uma interpretação satisfatória, de modo que consideramos que seja um ponto a ser investigado futuramente.

É importante, também, reconhecer que houve uma diferença no traço de animacidade do sujeito da oração principal no contraste entre os tipos de configuração. Os estímulos na versão alinhada apresentavam sujeito [+ animado] da oração principal (p.ex. **O engenheiro** *construiu os prédios que desabaram*), mas na versão aninhada o sujeito era [- animado] (p.ex., **O prédio** *que os engenheiros construíram desabou*). Desse modo, os resultados encontrados poderiam ser interpretados alternativamente como sendo decorrentes do fator animacidade do sujeito da oração principal, que não foi controlado.

Com o objetivo de investigarmos experimentalmente de forma mais precisa o papel da configuração das DNAs e controlar uma possível interferência da animacidade do sujeito, optamos por conduzir uma segunda versão do experimento, realizando algumas modificações nas sentenças experimentais apresentadas na primeira versão.

5. O processamento de múltiplas DNAs no PB – parte II

Estudos anteriores (DE VRIES *et al.*, 2012) sugerem que sentenças com dependências aninhadas causam custo maior no processamento porque, ainda que possuam a mesma quantidade de conteúdo, exibem limites severos na profundidade em que a informação é incorporada. No experimento descrito anteriormente, não foi possível captar claramente a relevância desse fator no processamento, já que o fator animacidade do sujeito não foi controlado. Buscamos, com esse novo experimento, superar as eventuais limitações existentes no anterior. Foi utilizada a mesma técnica experimental do experimento descrito anteriormente.

Design experimental e materiais: Novos estímulos foram criados a partir do que foi observado no experimento anterior. Os estímulos foram novamente distribuídos em quadrado latino gerado pelo próprio *software Linger*, que foi também utilizado na programação e aplicação desse experimento. Assim como no primeiro experimento, o experimento II foi formado por um conjunto de 12 sentenças experimentais e um conjunto de 20 sentenças distratoras. A principal modificação realizada na construção dos estímulos foi optar por sentenças que, ainda que estivessem na versão aninhada ficariam com o sujeito animado. A seguir apresentamos exemplos de cada condição experimental:

Quadro 4 – Condições experimentais.

NÚMERO DO SUJEITO	CONFIGURAÇÃO DA DNA	
	ALINHADA	ANINHADA
SINGULAR	O gato perseguiu <i>os ratos</i> que <i>entraram</i> no buraco.	O rato que <i>os gatos</i> <i>perseguiram</i> entrou no buraco.
PLURAL	Os gatos perseguiram <i>o rato</i> que <i>entrou</i> no buraco.	Os ratos que <i>o gato</i> <i>perseguiu</i> entraram no buraco.

Fonte: elaboração própria

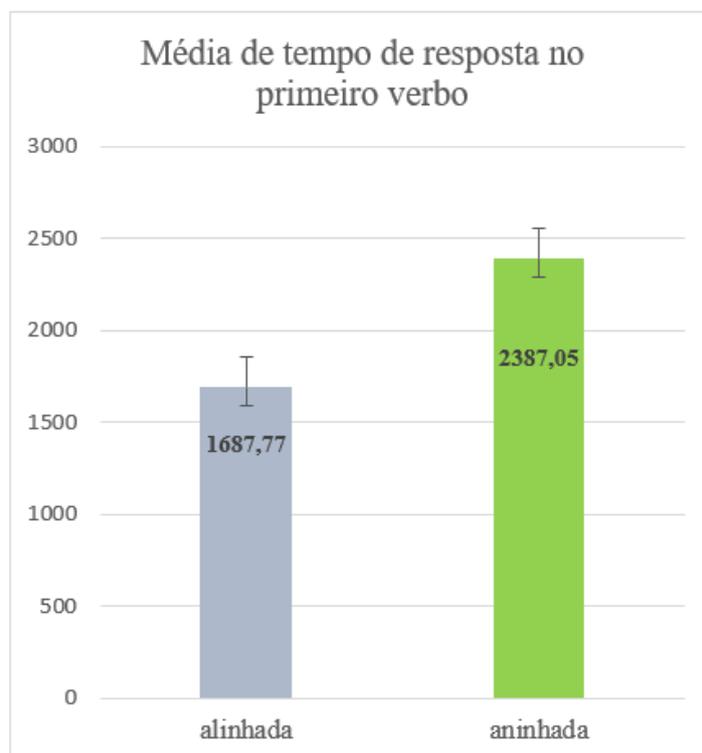
Previsões: Nossas previsões foram as mesmas do primeiro experimento.

Participantes: participaram voluntariamente 24 estudantes universitários de diferentes cursos (Serviço Social, Jornalismo, Letras, Enfermagem, Medicina, Bacharelado Interdisciplinar, Nutrição e Farmácia), com idade entre 18 e 31 anos (idade média: 21,1). Nenhum participante ou dado foi excluído.

5.1 Resultados e discussão

Os dados obtidos nesse segundo experimento foram submetidos ao teste de igualdade de variâncias do erro de *Levene* no *software Statistical Package for the Social Sciences* (SPSS) e revelaram homogeneidade da amostra em ambos os verbos. Os dados de tempos de escolha/leitura de cada segmento foram submetidos a uma análise de variância (ANOVA) de design fatorial 2x2 de medidas repetidas (*within-subjects*). No que diz respeito ao tempo de resposta no primeiro verbo, foi obtido efeito principal de *configuração de DNA* ($F(1,47) = 15,5$ $p < 0,000269$), com médias maiores nas condições com DNA aninhada. Nessa versão do experimento também não houve efeito principal da variável *número do sujeito*, assim como não foi encontrado efeito de interação. Foram realizados testes *post hoc* para comparações entre pares das condições (teste-t pareado unicaudal), que revelaram diferença estatisticamente significativa entre as condições DNA alinhada ($M=1687,77$ ms, $DevP=687,09$ ms) vs. DNA aninhada ($M=2387,05$ ms, $DevP=1167,425$ ms; $t(23)=5.4010$, $p < 0,0000$). O gráfico abaixo apresenta as médias e ilustra as diferenças no tempo de reação da escolha do verbo nas condições alinhada e aninhada independentemente do número do sujeito:

Gráfico 4: média tempo de reação no primeiro verbo



Fonte: elaboração própria

O quadro abaixo especifica as médias do tempo de reação do primeiro verbo em cada condição experimental:

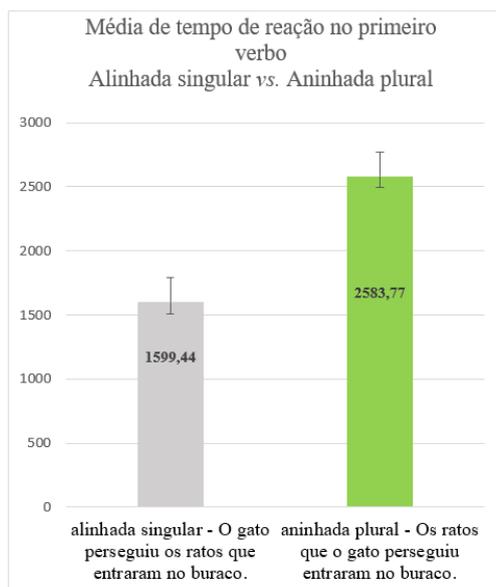
Quadro 5: média do tempo de reação no primeiro verbo em todas as condições experimentais

NÚMERO DO SUJEITO	Médias de tempo de escuta em cada condição – primeiro verbo	
	ALINHADA	ANINHADA
SINGULAR	1599,44ms	2190,33ms
PLURAL	1776,70ms	2583,77ms

Fonte: elaboração própria

Ainda no que diz respeito ao tempo de escolha do primeiro verbo, houve diferença significativa entre as condições *alinhada singular* (M=1599,44ms, DevP=633,08ms, EP=91,38ms) e *aninhada plural* (M=2583,77ms, DevP= 1313,49ms, EP=189,59ms; $t(47) = 4,15, p < 0,0001$). O gráfico abaixo mostra a diferença entre o tempo das condições comparadas:

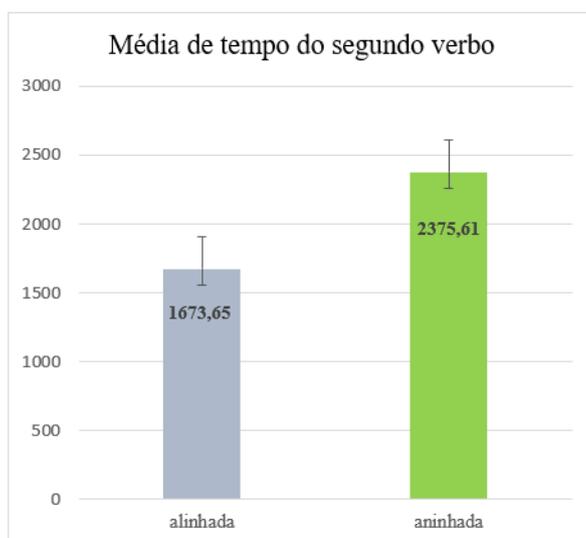
Gráfico 5: média do tempo de reação no primeiro verbo das condições alinhada singular e aninhada plural



Fonte: elaboração própria

No que concerne ao tempo de reação do segundo verbo, também foi registrado efeito principal de *configuração de DNA* ($F(1,47) = 9,56$ $p < 0,003348$), com médias maiores nas condições com DNAs em uma configuração aninhada. Foram realizados testes *post hoc* para comparação entre pares das condições (test-t pareado unicaudal), que ratificaram a diferença estatisticamente significativa entre as condições DNA alinhada ($M = 1673,635$ ms, $DevP = 812,77$ ms) vs. DNA aninhada ($M = 2375,615$ ms, $DevP = 1651,915$ ms; $t(23) = 3.47$, $p = 0,001$). O gráfico abaixo mostra a diferença de tempo entre as condições levando em consideração a variável *configuração de DNA*:

Gráfico 6: tempo de reação no segundo verbo



Fonte: elaboração própria

No quadro abaixo reportamos as médias de tempo de reação no segundo verbo em todas as condições experimentais:

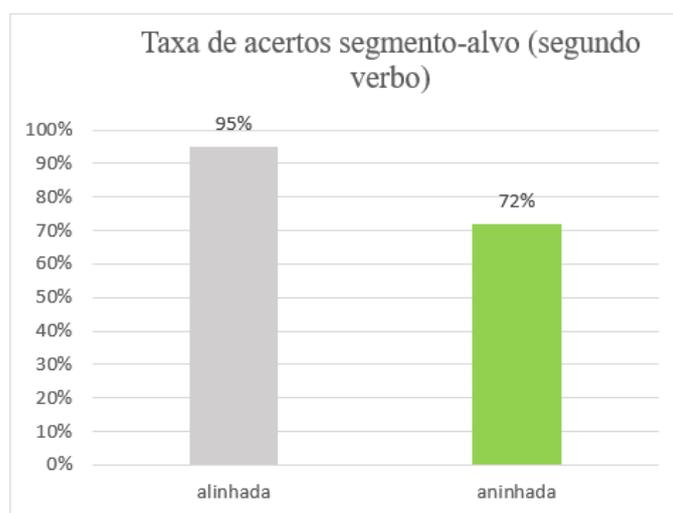
Quadro 6: média do tempo de reação no segundo verbo em todas as condições experimentais

NÚMERO DO SUJEITO	Médias de tempo de escuta em cada condição – segundo verbo	
	ALINHADA	ANINHADA
SINGULAR	1494,23ms	2557,31ms
PLURAL	1853,04ms	2193,92ms

Fonte: elaboração própria

No que tange à segunda variável dependente – taxa de acertos do segmento alvo – foram realizados testes estatísticos não paramétricos. O primeiro teste realizado foi o *Q de Cochran* e, posteriormente, o teste *McNemar*. No primeiro verbo não houve resultado estatisticamente significativo ($p < 0,416$). Os valores das taxas de acertos foram 100% na condição alinhada plural, 97% na condição alinhada singular, 97% na condição aninhada plural e 95% na condição aninhada singular. Entretanto, no segundo verbo, houve efeito principal de tipo de DNA, com taxa de acerto significativamente menor na condição aninhada, independentemente do número do verbo (Q de Cochran = 26,092^a = $p < 0,0001$):

Gráfico 7: taxa de acertos do segmento alvo (segundo verbo)



Fonte: elaboração própria

Os valores obtidos em cada condição foram de 93% e 97% nas condições alinhada plural e alinhada singular, respectivamente. As condições aninhadas registraram total de 72% para a condição plural e 72% na versão com número singular.

O resultado das análises das condições sugere que a condição aninhada, independentemente

do número do sujeito, tem o custo de processamento maior, refletido em tempos maiores de reação, tanto no primeiro verbo, quanto no segundo verbo. Os resultados relativos ao segundo verbo são particularmente interessantes pois, apesar de os elementos dependentes estarem próximos, houve dificuldade na recuperação do número do sujeito, indicando que a sobreposição de dependências pode causar sobrecarga na memória de trabalho. As taxas de acerto do segmento alvo, menores nas condições aninhadas independentemente da variável número, também apontam nessa direção. Neste segundo experimento, controlamos a animacidade do sujeito. Tal fator não havia sido controlado no primeiro experimento e poderia ser uma interpretação alternativa para explicar seus resultados. No entanto, mesmo controlando a animacidade do sujeito, obtivemos efeito de configuração de DNAs, que parece ter sido, portanto, o fator relevante para o processamento dos estímulos.

Considerações finais

Neste artigo buscamos discutir o efeito que as diferentes configurações de múltiplas DNAs poderiam exercer no processamento da concordância em uma língua natural. Para isso, elegemos relações de concordância verbal em orações principais e relativas no PB, que permitem contrastar dois tipos de configuração de DNAs encontrados nessa língua: alinhada, em que uma DNA se segue a outra; e aninhada, em que uma DNA se encaixa – se aninha – entre os elementos de outra DNA. Realizamos dois experimentos com adultos falantes nativos do PB através da tarefa de labirinto, tendo como variáveis independentes o tipo de configuração e número; e variáveis dependentes tempo de resposta (medido nos primeiro e segundo verbos) e taxa de acerto da escolha dos verbos. Os estímulos do primeiro experimento apresentavam variação do traço de animacidade do sujeito, ora [+animado], ora [-animado], o que foi neutralizado no segundo experimento, apenas com sujeitos [+animado].

Em conjunto, os resultados dos experimentos foram, de modo geral, ao encontro de nossas previsões e sugerem que a configuração das DNAs interfere no custo do processamento das sentenças. Foi encontrado efeito principal de configuração de DNA no tempo de resposta do primeiro verbo nos dois experimentos, e não houve efeito principal de número nem interação entre os fatores. No segundo experimento, buscamos neutralizar o efeito da animacidade do sujeito entre as condições aninhadas e alinhadas, que poderia ter sido um fator relevante mascarando o efeito da configuração de DNA. No entanto, o segundo experimento também apresentou efeito principal de tipo de configuração no segundo verbo, afastando tal interpretação. Ainda, foram observadas diferenças estatisticamente significativas no tempo de resposta ao primeiro verbo entre pares que contrastavam o tipo de configuração (alinhada vs. aninhada), mas apresentavam o mesmo número (plural no primeiro experimento; singular, no segundo) e estavam próximos linearmente. Esses resultados sugerem fortemente que a complexidade

de processamento é do tipo de configuração das dependências não adjacentes que, quando inseridas no meio de outra geram custo maior, e não da distância linear entre os elementos codependentes ou do número do sujeito. A taxa de acertos do verbo-alvo, menor no segundo verbo da condição aninhada nos dois experimentos, reforça essa ideia. Assim, apesar de haver a mesma quantidade de informação nas sentenças alinhadas e aninhadas, o nível em que a informação é incorporada nas sentenças aninhadas é mais profundo, havendo de certa forma sobreposição de dependências.

Na breve revisão sobre processamento de orações relativas, apresentada na seção 3, pudemos observar que a ordem e a distância entre os elementos correlacionados, assim como a animacidade do sujeito, são apontados como fatores responsáveis por maior ou menor dificuldade de processamento dessas estruturas. Nossos resultados são parcialmente compatíveis com as propostas desenvolvidas, já que apenas a ordem – mais apropriadamente, a organização, a configuração – das múltiplas DNAs teve influência no processamento. Nesse caso, parece haver diferença no processamento entre distância estrutural e distância linear.

No limite de nosso conhecimento, este é primeiro estudo que trata de processamento adulto de diferentes configurações de múltiplas DNAs em uma língua natural a partir de orações relativas. Assim, buscamos lançar luz sobre fatores que podem estar envolvidos nesse processamento e contribuir para a abertura de uma frente promissora de pesquisa e uma melhor compreensão do fenômeno em questão.

REFERÊNCIAS

BEVER, Thomas. The cognitive basis for linguistic structures. In: HAYNES, John. (org.) *Cognition and development of language*. New York: Wiley, p. 279 – 362, 1970

DE VRIES, Meinou., CHRISTIANSEN, Morten., PETERSSON, Karl Magnus. Learning recursion: multiple nested and crossed dependencies. *Biolinguistics*, v. 5, p. 10 – 35, 2011.

DE VRIES, Meinou, PETERSON, Karl Magnus, GEUKES, Sebastian., ZWITSERLOOD, Piene., CHRISTIANSEN, Morten. Processing multiple non-adjacent dependencies: evidence from sequence learning. *Philosophical transactions of the Royal Society B: Biological Sciences*, p. 2065-2076, 2012.

FIEBACH, Christian.; SHLESEWSKY, Matthias.; FRIEDERICI, Angela. Separating memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, v. 47, p. 250-272, 2002.

FRAZIER, Lyn.; FODOR, Janet Dean. The sausage machine: A new two-stage parsing model. *Cognition*, v. 6, p. 291-325, 1978.

GIBSON, Edward. Linguistic complexity: locality of syntactic dependencies. *Cognition*, v. 68, p. 1-76, 1998.

GÓMEZ, Rebecca. Variability and detection of invariant structure. *Science* v.13, p.431-36, 2002.

GÓMEZ, Rebecca.; MAYE, Jessica. The developmental trajectory of nonadjacent dependency learning. *Infancy*, v. 7(2), p.183-206, 2005.

GORDON, Peter.; HENDRICK, Randall.; JOHNSON, Marcus. Memory Interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, v. 27, p. 1411-1423, 2001.

HENRIQUE, Késsia. *Variação linguística e processamento: investigando o papel da distância entre sujeito e verbo na realização da concordância verbal variável no PB*. 2016. 149 f. (Dissertação de mestrado) – Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2016.

LAGUARDIA, Milene. *Etapas iniciais da aquisição lexical – Habilidades estatísticas e simbólicas no tratamento de dependências não adjacentes*. 2016. 124 f. (Tese de doutorado) - Faculdade de Letras, Universidade Federal de Juiz de Fora, Juiz de Fora, 2016.

LAGUARDIA, Milene.; NAME, Cristina.; SHI, Rushen. Sensitivity to non-adjacent dependencies and grammatical category relations in 11 month-old infants. *BUCLD* 38, 2013.

LAGUARDIA, Milene.; SANTOS, Erica.; SHI, Rushen.; NAME, Cristina. Eleven-month-old infants use prosodic boundaries to learn non-adjacent grammatical dependencies. *BUCLD* 40, 2015.

MARCUS, Gary.; VIJAYAN, S.; BANDI RAO, Shoba.; VISHTON, Peter. Rule learning by seven-month-old infants. *Science*, v. 283, p.77-80, 1999.

MCWHINNEY, Brian. The competition Model. In B. MacWhinney (ed.), *Mechanisms of language Acquisition*. Hillsdale, NJ: Erlbaum, p. 249-308, 1987.

MCWHINNEY, Brian.; PLEH, Csaba. The Processing of Restrictive Relative Clause in Hungarian. *Cognition*, v. 29, p. 95-141, 1988.

MIRANDA, Fernanda. *O Custo de processamento de orações relativas: um estudo experimental sobre relativas com pronome resumptivo no Português Brasileiro*. 2008. 170 f. (Dissertação de Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

NEWPORT, Elisa.; ASLIN, Richard. Learning at a distance: I. Statistical learning of nonadjacent dependencies. *Cognitive Psychology*, v. 48, p. 127– 162, 2004.

NAME, Cristina.; LAGUARDIA, Milene. Etapas iniciais da aquisição lexical. *Prolíngua*, 8(2), p. 296-308, 2013.

NAME, Cristina.; SHI, Rushen; KOULAGUINA, Elena. 2011. Non-adjacent dependency learning and grammatical categorization in 11- month-old infants. *12th International Congress for the Study of Child Language*. 2011.

PRITCHETT, Bradley. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press, 1992.

SANDOVAL, Michelle.; GÓMEZ, Rebecca. The development of nonadjacent dependency learning in natural and artificial languages. *WIREs Cogn Sci*. doi: 10.1002/wcs. 1244, 2013.

SANTELMANN, Lynn.; JASCZYK, Peter. Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition*, v. 69, p. 105-134, 1998.

SHELDON, Amy. A role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, v.13, p. 272-281, 1974.

TOWNSEND, David. J.; BEVER, Thomas. *Sentence comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press, 2001.

TRAXLER, Mattew. MORRIS, Robin., SEELY, Rachel. Processing Subject and object relative clauses: evidence from eye movements. *Journal of Memory and Language*, v. 47, p. 69-90, 2002.

VAN DEN BOS, Esther.; CHRISTIANSEN, Morten.; MISYAK, Jennifer. Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. *Journal of Memory and Language*, 67, 507-520, 2012.

ZUHURUDEEN, Fathima.; HUANG, Yi-Ting. Effects of statistical learning on the acquisition of grammatical categories through Qur’anic memorization: A natural experiment. *Cognition*, 148, 79-84, 2016.

**O LABORATÓRIO NA ESCOLA E O DESENVOLVIMENTO DA METACOGNIÇÃO
LINGUÍSTICA: UMA EXPERIÊNCIA TRANSLACIONAL COM LEITURA**
*THE LABORATORY IN SCHOOL AND THE DEVELOPMENT OF LINGUISTIC
METACOGNITION: A TRANSLATIONAL EXPERIENCE WITH READING*

Erica dos Santos Rodrigues¹

RESENHA DE:

MAIA, Marcus (Org.). Psicolinguística e Metacognição na Escola. Campinas, SP: Mercado das Letras, 2019. ISBN 978-85-7591-589-9

Os resultados de avaliações em larga escala de alunos da Educação Básica no Brasil apontam para uma situação dramática, com graves implicações econômicas, sociais e pessoais. Dados de 2018 do PISA, sigla em inglês para o Programa Internacional de Avaliação de Estudantes, revelam que, na última década, praticamente não houve mudanças em relação ao desenvolvimento dos alunos em leitura.² Uma análise do percentual de estudantes por nível de proficiência em Leitura no PISA desse ano revela que, embora 50% dos estudantes brasileiros tenham alcançado o “nível 2 ou acima” em letramento em Leitura – sendo o nível 2 considerado pela *Organização para Cooperação e Desenvolvimento Econômico* como nível mínimo de proficiência que as crianças devem adquirir até o final do Ensino Médio (OECD, 2019) –, os outros 50% ainda permanecem no nível 1, ou seja, têm dificuldades em ler textos de tamanho e complexidade moderadas, sobre tópicos não familiares, e precisam de tarefas com dicas explícitas sobre “o que precisa ser feito, como fazê-lo e onde, no(s)

1 Profa. do Programa de Pós-Graduação *Estudos da Linguagem*, PUC-Rio; pesquisadora do LAPAL- Laboratório de Psicolinguística e Aquisição da Linguagem. Bolsista de Produtividade do CNPq/PQ2.

Email: ericasr@puc-rio.br

2 O PISA é uma iniciativa da Organização para Cooperação e Desenvolvimento Econômico (Organisation for Economic Co-Operation and Development – OECD, em inglês), com realização a cada três anos. O Programa avalia três áreas de conhecimento - leitura, matemática e ciências, com foco em um desses domínios em cada edição. O Brasil participa das avaliações desde seu início em 2000. Na última edição, em 2018, a avaliação atingiu no país mais de 600 mil estudantes de 15 anos em 79 países ou regiões diferentes. Informações detalhadas sobre o PISA podem ser obtidas na página do INEP, órgão responsável pelo planejamento e condução do PISA no Brasil - <http://inep.gov.br/pisa>

texto(s), os leitores devem concentrar sua atenção.”

Vários fatores, de naturezas distintas, podem ser elencados para explicar essa situação. Em termos acadêmicos/científicos, uma pergunta que se coloca é o quanto evidências robustas de pesquisas sobre o desenvolvimento de estratégias cognitivas e metacognitivas dos alunos têm sido consideradas e, de algum modo, alimentado a reflexão sobre o trabalho com leitura em sala de aula.

É nesse cenário que o livro *Psicolinguística e Metacognição na escola*, publicado em 2019, pela Editora Mercado de Letras, se apresenta como resultado de um trabalho que provê informações fundamentais sobre a dimensão cognitiva da leitura. Organizado pelo prof. Marcus Maia, da UFRJ, a obra é um dos frutos de um projeto iniciado em 2017, com duração de dois anos, que reuniu professores, pesquisadores e alunos da área de Psicolinguística de três instituições do Rio de Janeiro – UFRJ, UERJ e CEFET-RJ, reunidos no Laboratório LER - Laboratório de Eletrofisiologia e Rastreamento Ocular (<http://www.ler.lettras.ufrj.br/>). Diferentemente de outras obras sobre leitura, nesse livro reportam-se resultados de pesquisas acerca do processo de leitura em tempo real, sobre estratégias de leitura, relacionando dados obtidos por ferramentas como o rastreador ocular e a eletrofisiologia (EEG-ERP), além de paradigmas clássicos, como o teste de Cloze. Os participantes são alunos das séries finais do Ensino Fundamental (EF) de uma escola pública, da rede municipal do Rio de Janeiro, o Colégio Estadual Joaquim Távora (CEJOTA), localizado em Niterói, RJ.

O livro está estruturado em oito capítulos e conta com prefácio da profa. Eloisa Pilati (UNB) e apresentação do organizador. Todos os capítulos são escritos pelos participantes do projeto, com registro das atividades realizadas no próprio “chão da fábrica”, isto é, no CEJOTA, numa busca de estabelecer pontes, nexos, entre a academia e a escola, o que dá a dimensão translacional da pesquisa realizada. Como será possível acompanhar na apresentação dos capítulos, o livro reúne (i) resultados de testes realizados no primeiro semestre de 2017, na fase inicial de diagnóstico do projeto, em que se fez uso de técnicas e paradigmas distintos – como a técnica de rastreamento ocular, eletrofisiologia, leitura automonitorada e de questionários baseados no paradigma de Cloze; (ii) relatos de oficinas que trabalharam tópicos específicos – como a leitura de estruturas ambíguas e a expressão de perspectiva; e (iii) resultados de retestagem em que foi possível comparar o desempenho dos alunos do EF que fizeram as oficinas com alunos do ensino superior (ES) e um grupo controle do EF, que não participou das oficinas.

Desafiando paradigmas pedagógicos e mesmo algumas abordagens na área de leitura, o projeto assume o período como a unidade mais relevante para se diagnosticar a competência leitora

inicial e para se desenvolverem práticas de natureza metacognitiva, que impulsionem o raciocínio científico, na linha do que propõem Honda e O’Neil (1993). O período é a unidade mínima, a unidade estruturante do texto, e, nesse sentido, analisar o comportamento leitor nesse nível pode ser informativo sobre processos básicos, necessários à compreensão global do texto. Em termos teóricos, há um alinhamento com a Teoria da Gerativa e com achados da área de Psicolinguística, em especial a área de processamento de frases e processamento de palavras.

Leitores mais diretamente vinculados à Academia bem como aqueles com atuação direta na escola e os que estão se preparando para isso, os licenciandos, terão nesse livro a oportunidade de experimentar uma nova perspectiva no trabalho com a língua. Poderão ver como é possível conectar ciência e educação, e como esta linha de trabalho pode contribuir para o desenvolvimento da metacognição, tão essencial ao desenvolvimento, não só do raciocínio científico, mas à formação de cidadãos críticos e questionadores.

O capítulo de abertura intitulado “Metacognição e Educação Linguística” é o fio-condutor, que traz a perspectiva teórico-metodológica que está na base de todo o projeto. Escrito por Marcus Maia, Daniela Cid de Garcia e Mariana Fernandes, o capítulo introduz o conceito de metacognição e apresenta o modelo de monitoramento cognitivo (FLAVELL, 1976), que busca integrar, de modo dinâmico, conhecimento e experiência metacognitiva, objetivos e estratégias. Com base em Pilati (2017), os autores observam que grande parte das atividades gramaticais realizadas na escola não promove reflexão e tem pouca conexão com a vida prática. Nesse contexto, discutem o caráter plural do conceito de língua – como fenômeno social e cognitivo, e sinalizam para a ausência de perspectivas com orientação formalista, em especial a da Gramática Gerativa, na formação dos professores. Maia, Garcia e Fernandes buscam explicitar a contribuição que o estudo formal da linguagem pode representar para o desenvolvimento da literacia plena, deixando claro, não obstante, que a transposição dessa visão para o ensino não implica uma aplicação direta dos achados teóricos. Trata-se, nos termos dos autores, ancorados em Honda e O’Neil (1993), “de utilizar métodos e achados sobre a linguagem para incentivar o uso do pensamento racional e científico” (p. 28)”. Com base na noção de inatismo, falam da importância de as práticas pedagógicas reconhecerem o conhecimento linguístico complexo que a criança já traz ao ingressar na educação formal. Cabe à escola desenvolver a consciência metalinguística do aluno e ampliar seu repertório linguístico. A partir deste ponto, os autores apresentam as bases do projeto realizado no CEJOTA. Justificam a definição do período como a unidade escolhida para os estudos e para o trabalho nas oficinas. Para os autores, o período é o *building block*, a unidade estrutural fundamental, que pode permitir o desenvolvimento da metacognição da linguagem e pode atuar como propulsor da competência

leitora. Em seguida, introduzem brevemente a técnica de rastreamento ocular, a qual será detalhada nos capítulos III e IV, buscando apontar o tipo de inovação que esta espécie de microscópio da leitura pode trazer para a metacognição da linguagem.

O capítulo II “Educação e trabalho: concepção politécnica na formação de leitores”, de autoria de Sabrina Santos, Paulo Maia, Thiago Albuquerque e Rafaela Aquino, apresenta o relato de como foram conduzidas duas oficinas de cunho híbrido – linguístico e artístico - viabilizadas a partir de uma parceria entre o LAPEX³, coordenado pelo prof. Marcus Maia, e o Laboratório de extensão Grupo de Educação Multimídia (GEM), coordenado pelo prof. Paulo Maia (UFRJ). As oficinas foram realizadas com alunos do 8º ano do EF do CEJOTA e tinham como objetivo tornar salientes para esses alunos mecanismos formais relacionados a efeitos de perspectiva e de ambiguidade na interpretação da linguagem. No que tange à expressão de perspectiva, a proposta das oficinas partiu dos achados de Ribeiro (2017) e de Maia (2017, 2018), que, em conjunto, mostram diferenças de comportamento ocular entre alunos do EF e do Ensino Superior (ES, estudantes do curso de Letras da UFRJ) na leitura de períodos complexos⁴. Em relação ao processamento de estruturas ambíguas, tomou-se como base resultados do estudo realizado por Silvano (2018), segundo o qual alunos do EF tendem a realizar um processamento superficial, não algorítmico de estruturas ambíguas, chamado processamento *good-enough*, o que pode levar a falhas na interpretação final das sentenças. Em termos de metodologia, as oficinas se estruturaram a partir do entrelaçamento da concepção de trabalho como princípio educativo, que está na base da metodologia politécnica (SAVIANI, 2003) com a visão de Honda e O’Neil (1993; 2017), segundo a qual o conhecimento linguístico pode ser um instrumento de formação do pensamento científico. Esse entrelaçamento foi construído a partir de atividades pedagógicas de tradução intersemiótica, as quais resultaram na produção de uma revista no formato fan(zine) com base no capítulo “Baleia”, do livro *Vidas Secas*, de Graciliano Ramos, e de uma cena do filme homônimo de Nelson Pereira dos Santos e na produção de fitas para praxinoscópio, a partir do conto “O Outro”, de Rubem Fonseca, e da análise de estudos de fisiologia de movimentos do corpo humano, de E. J. Muybridge, no séc. XIX. No processo de produção dos materiais das oficinas, foi possível levar os alunos a refletir sobre as particularidades de cada linguagem e a operar com as noções de ambiguidade e de ponto de vista/perspectiva, promovendo, assim, raciocínio metalinguístico. Usando as palavras dos autores, pode-se dizer que o trabalho produtivo realizado permitiu levar “a

3 “O grupo de pesquisa LAPEX, da UFRJ/CNPq, estuda a estrutura morfo-sintática e os processos de parsing e de interpretação na compreensão e produção de frases e palavras em línguas naturais, por parte de sujeitos normais ou com distúrbios linguísticos. O grupo realiza pesquisa de natureza teórica e experimental, utilizando diferentes protocolos psicolinguísticos, tais como os paradigmas de rastreamento ocular, priming, leitura auto-monitorada, julgamento imediato de gramaticalidade, decisão lexical e outros.” (<https://lapex.wordpress.com> acessado em 9 de maio 2020)

4 O estudo de Maia (2018) será detalhado na resenha do capítulo III.

lógica laboratorial para a sala de aula atuando sobre a indissociabilidade entre teoria e prática” (p. 62). O relato detalhado da condução das oficinas deixa claro como o estudo da língua pode ser feito em outras bases – que não a memorização de regras e de terminologia - e motiva o leitor a considerar como o estudo da língua pode ser um instrumento para o desenvolvimento da metacognição.

O capítulo III, intitulado “Rastreamento ocular de períodos compostos e consciência sintática”, elaborado pelo prof. Marcus Maia, apresenta os resultados do estudo e das oficinas de rastreamento ocular da leitura de períodos compostos por subordinação, realizados durante os dois anos do projeto Linguística para a Educação Básica. Nesse capítulo, o autor reforça as considerações feitas no capítulo I sobre a relevância do período no diagnóstico da compreensão leitora inicial e no desenvolvimento da metacognição. Acrescenta ainda que é um construto sobre o qual já há muita pesquisa – tanto no âmbito da Teoria Gerativa como na área de Processamento de Frases, na Psicolinguística - mas cuja relevância para a área de leitura tem sido negligenciada. O capítulo se organiza em três seções principais, as quais correspondem às três fases do trabalho realizado no CEJOTA: a fase de testagem inicial, que consistiu num experimento de rastreamento ocular aplicado no primeiro semestre do projeto; a realização de oficinas, que foram planejadas a partir dos resultados da testagem inicial; e a fase de retestagem, com reaplicação do experimento de rastreamento ocular para aferir os resultados das oficinas. No experimento realizado, originalmente apresentado em Maia (2018), o autor parte da hipótese de que, nos períodos construídos para o estudo, a oração principal (OP) deve ser computada com maior atenção na leitura, por apresentar maior completude semântica e ser responsável pela expressão do ponto de vista do período do que a subordinada. Esperam-se, pois, no caso de leitores proficientes, maiores latências de fixação ocular associadas à OP do que às subordinadas. O estudo envolveu a leitura de períodos subordinativos formados por quatro orações, seguidas de pergunta interpretativa dirigida à oração principal, e foi aplicada a alunos do ES e do 8º ano do EF. O resultado obtido foi na direção esperada: enquanto os alunos do ES leem o período completo e, na segunda passagem do olhar, buscam identificar o seu centro nuclear, os alunos do EF fazem uma leitura linear, privilegiando a ordem de menção e não a estrutura hierárquica. A proposta das oficinas com os alunos do EF foi o de buscar reverter esse padrão ocular, de modo a que estes pudessem realizar uma análise estruturante e, assim, vir a extrair a informação central do período, referida por Maia como o *gist* do período. As oficinas, que tiveram duração de dois semestres, apresentaram uma multiplicidade de dinâmicas, e fizeram uso de materiais diversos – verbais e não verbais, o que permitiu uma conceituação ampla da ideia de perspectiva e ponto de vista. No capítulo III, como já dito, Maia se detém na apresentação das oficinas que fizeram uso de dados de rastreamento ocular dos próprios alunos como estratégia para desenvolvimento da metacognição. Em consonância com a proposta

de Mason, Pluchino e Tornatora (2016), a técnica foi utilizada como ferramenta educacional. O interessante das oficinas é que os alunos aprenderam a “ler” as imagens geradas pelo rastreador – mapas de calor, *gazeplots* estáticos ou dinâmicos - e a relacioná-las às respostas dadas às perguntas interpretativas, tudo conduzido de forma lúdica, buscando estimular o olhar científico. A fase 3, da retestagem envolveu experimento similar ao da fase de teste, e foi aplicado a alunos que fizeram a oficina, a alunos do ES e a um grupo controle de alunos do 9º ano que não participaram das oficinas. Os resultados, tanto relativo às medidas *on-line* (dados oculares) quanto à medida *off-line* (perguntas de interpretação), indicam que os alunos que participaram das oficinas aproximam-se dos alunos do ES e distanciam-se do grupo controle. A atenção direcionada à oração principal é evidenciada pelas maiores latências de fixação nessa oração, na diferença significativa de percentual global de fixações entre as OPs e as subordinadas na primeira passagem do olhar e também no índice de *lookbacks*, tendo os alunos do 8º ano realizado um percentual bastante alto (80%) de regressões em todas as posições de OP. Os resultados relativos à medida *off-line* também evidenciaram diferenças entre o grupo de EF que fez a oficina e o grupo controle. Esses resultados indicam que o objetivo translacional das oficinas foi alcançado e que o rastreamento ocular pode ser utilizado como uma espécie de microscópio metacognitivo, usando os termos de Maia. Cumpre, não obstante, notar que o mesmo grupo de alunos também participou de outras oficinas, como as reportadas no capítulo II e V do livro; logo, o desenvolvimento da competência leitora dos alunos do EF precisa ser avaliada em função do conjunto de atividades realizadas.

Escrito por Emily Silvano e Marcus Maia, o capítulo IV “Leitura de períodos temporariamente ambíguos: uma proposta de estudo psicolinguístico translacional para a educação básica” apresenta os resultados de estudo e oficinas centrados no fenômeno da ambiguidade estrutural. São examinadas sentença ambíguas do tipo “A aluna disse à professora que chegou tarde que precisava sair mais cedo”. Nesse tipo de sentença, o processador sintático, com funcionamento automático e guiado por princípios de economia, tende a concatenar a oração sublinhada em destaque ao verbo (“disse”), ou seja, privilegia uma análise da oração como complemento verbal. No entanto, essa análise mostra-se insatisfatória em relação à continuidade da sentença, o que provoca um efeito que vem sendo explicado na literatura psicolinguística pela Teoria do *Garden-path* (FRAZIER; FODOR, 1978): nessas sentenças, de modo similar ao que se observa quando um indivíduo está perdido em um labirinto (*garden-path*), é necessário, para poder sair do labirinto, abandonar uma caminho inicialmente escolhido. Ocorre uma reanálise da estrutura – no exemplo dado, a oração sublinhada é concatenada ao substantivo “professora”, assumindo função de oração adjetiva. Assim como no estudo e nas oficinas apresentadas no capítulo V, aqui também se parte da ideia de que a linguística

e a psicolinguística, com seus métodos e técnicas experimentais, podem atuar como ferramentas educacionais para o desenvolvimento do raciocínio científico. Os autores, nesse capítulo, ampliam a discussão sobre o pensar linguisticamente e seus efeitos, e também exploram a potencialidade da técnica de rastreamento ocular como microscópio metacognitivo. De forma paralela ao trabalho com períodos compostos por subordinação, registram-se as três fases do projeto para o trabalho com ambiguidades: testagem inicial, oficinas e retestagem. Na primeira testagem, no semestre de diagnóstico, participaram alunos da etapa final do EF e alunos de ES da UFRJ. A fase de retestagem foi realizada com alunos do EF que fizeram as oficinas, grupo controle de alunos do EF da mesma escola e alunos de ES, estudantes da UFRJ. O experimento de rastreamento ocular consistia na leitura de frases com e sem ambiguidade estrutural. Esperava-se um efeito principal associado à ambiguidade, com tempos totais de fixação ocular maiores para as sentenças ambíguas, em função da reanálise estrutural, e também menor número de acertos em pergunta de compreensão. Em relação aos grupos, a previsão para o teste inicial era de que os alunos do EF apresentassem tempo maior de leitura, e que, na retestagem, houvesse redução desse tempo para os alunos que participassem das oficinas. Em direção contrária ao esperado com base na literatura, no teste inicial, os alunos do EF apresentaram tempos de leitura e quantidades de fixação mais baixos do que os alunos mais velhos, o que é interpretado como resultado de um desengajamento da tarefa experimental. Além disso, apresentaram maior número de erros do que os alunos do ES. Quanto às oficinas, foi realizada uma série de três: a primeira consistia em um jogo de *Role Playing Game* (RPG), que envolvia resolução de ambiguidades para que o jogador pudesse avançar no tabuleiro; uma segunda oficina envolvendo análise de dados oculares; e uma terceira oficina, com elaboração de sentenças ambíguas para jogo de RPG e previsão de padrão de fixação e movimentação sacádica. Cabe ressaltar que o projeto sempre buscou integrar a dimensão lúdica e criativa, à reflexão sobre a linguagem e ao estímulo ao pensamento científico. Em relação ao resultado das retestagens, manteve-se o efeito de estrutura, como esperado, e obteve-se efeito de grupo – os alunos do ES foram os que leram mais rapidamente os dois tipos de sentença, e o grupo das oficinas, apesar de diferir do ES, atingiu a compreensão mais rapidamente do que o grupo controle. A mudança observada na retestagem, com aumento no tempo de leitura dos alunos do EF do grupo experimental, em especial nas sentenças *garden-path*, mostra maior engajamento dos participantes na tarefa. Os resultados relativos à tarefa interpretativa (*off-line*) indicaram melhor desempenho para os alunos do ES, seguido do grupo oficina e, por último, o grupo controle, que, a despeito ter tido mais um ano de escolaridade, manteve percentual muito elevado de erros nas sentenças *garden-path* – acima de 60%, o que é indicativo de um processamento superficial, *good-enough*, das sentenças testadas.

O capítulo V “Oficina do período: identificação, desmontagem e remontagem dos pontos de vista”, de Sara Ribeiro e Ana Luiza Machado, detalha a oficina que buscou exercitar metacognição da perspectiva, a partir de atividades de reflexão, desarticulação e reformulação de períodos subordinados. Ancoradas em Honda e O’Neil (2008), as autoras afirmam que a falta de consciência estrutural do período levaria os alunos a não considerarem a estrutura sintática como sendo diretamente relevante para a compreensão de frases e textos, determinando um desengajamento da leitura. O capítulo segue uma estrutura similar ao do capítulo II: inicia com reflexão teórica sobre o tópico das oficinas e, em seguida, relata como estas foram desenvolvidas. As autoras, na seção teórica, chamam a atenção para a ideia de que uma das características que difere os humanos dos outros seres é sua capacidade de diferenciar perspectivas e pontos de vista, isto é, os humanos apresentam teoria da mente (ASTINGTON; GOPNIK, 1991). Pontuam a importância de estimular o aluno a considerar perspectivas e pontos de vista implícitos, de modo a que possam pensar criticamente sobre as informações a que estão expostos. Citando Pilati (2017), observam também que a educação linguística que vem sendo conduzida na escola não prioriza essa reflexão e apresenta pouca ligação com a vida prática, o que pode gerar desmotivação no aluno. Ribeiro e Machado constroem as oficinas inspiradas em atividades de desmontagem estrutural e ideológica do período propostas por Maia (2006) no âmbito do ensino de língua portuguesa para alunos indígenas, e também orientadas por dados de rastreamento ocular de alunos do EF e do ES, no primeiro semestre do projeto, na etapa de diagnóstico. As oficinas foram realizadas semanalmente com uma das turmas do 9º ano, em tempos vagos. As atividades iniciais buscaram trabalhar a noção de perspectiva em material visual: imagens ambíguas ou que propiciassem a identificação de pontos de vista distintos. Após essa fase, foi realizada uma série de oficinas de montagem e desmontagem de orações em períodos, as quais foram precedidas por uma atividade de identificação de perspectiva com base na análise integrada de uma imagem e de um período de um texto de reportagem. Um ponto importante a destacar é que a metalinguagem usada na classificação das orações foi introduzida gradativamente, já ancorada na percepção de diferenças entre OPs e subordinadas. As oficinas trabalharam não apenas a compreensão das estruturas, mas também sua produção, com a criação de períodos que expressassem perspectivas diferentes para descrever o que imagens representavam. O capítulo é rico em exemplos do material utilizado nas oficinas, que sempre buscaram integrar informação verbal e não verbal, permitindo, pois, o desenvolvimento de habilidades de compreensão multimodal. Tanto este capítulo como o capítulo II são excelentes exemplos de como é possível e – eu diria, necessário – explorar metodologias ativas que permitam ao aluno acionar seu raciocínio científico de modo a desenvolver sua capacidade produtiva e de crítica.

Migrando do nível do período para o nível da palavra, o capítulo VI “Rastreamento ocular de palavras: uma comparação entre alunos do ensino fundamental e do ensino superior”, de autoria de Aline Saguie, Sabrina Santos e Daniela Cid de Garcia, busca promover um encontro entre os achados da área da Psicolinguística sobre processamento de palavras e o estudo da morfologia na escola. As autoras buscam aferir, a partir de estudo comparativo realizado com alunos do EF e do ES, como o conhecimento morfológico é utilizado no processo de leitura. O olhar microscópico recai sobre o morfema, tomado como unidade significativa. O objetivo do estudo é verificar o impacto dessa unidade no reconhecimento da palavra por via da leitura. A seção inicial do capítulo apresenta uma discussão sobre o tratamento dado à morfologia na gramática tradicional. Partindo das críticas de Silva e Medeiros (2016), discute-se, a partir de exemplificação, como é superficial e incoerente o tratamento que os fenômenos morfológicos recebem no âmbito da gramática normativa, a qual em geral se atém a terminologias e prescrições. Na seção seguinte, as autoras buscam mostrar a importância da identificação rápida de palavras no processo de leitura, entendida como um processo complexo que combina operações *bottom-up*, relativas ao reconhecimento ascendente da informação visual, com processos *top-down*, associados ao conhecimento prévio trazido pelo leitor. Em continuidade a estes tópicos, são mencionados os principais modelos de processamento de palavras e apresenta-se o recurso da transposição de letras – em que letras são deslocadas de sua posição original - como medida manipulativa para identificar fenômenos relacionados a propriedades morfológicas das palavras. Este recurso – aplicado no interior do morfema - é o empregado pelas autoras em experimento de rastreamento ocular, com tarefa *off-line* de decisão lexical, conduzido com alunos do 8º ano do EF e alunos do ES, com vistas a investigar se a composição morfológica de palavras multimorfêmicas (ex. multifacetado) é acessada pelo leitor e se os diferentes tipos de morfemas (prefixo, raiz, sufixo) possuem a mesma carga de processamento. Os resultados do experimento são bastante reveladores dos padrões de leitura de cada grupo: os alunos do ES fazem uma leitura mais eficiente e dedicam mais atenção ao morfema raiz, que carrega maior carga semântica. Isso é evidenciado pelo fato de a transposição intramorfêmica afetar mais as raízes tanto nos dados oculares (tempos médio de fixação mais elevados para este tipo de morfema em comparação ao prefixo e ao sufixo) quanto na tarefa de decisão lexical (a transposição de letras na raiz foi a que induziu mais erros de reconhecimento da palavra como pertencendo ao português brasileiro). Os alunos do EF, por sua vez, parecem ser guiados não pela estrutura, mas sim pela ordem de menção: os tempos médios de fixação são mais elevados no prefixo. Além disso, a taxa de erros desses alunos na tarefa *off-line* superou a do ES para todos os morfemas. Em suma, os resultados indicam que os leitores mais experientes fazem uso de seu conhecimento morfológico como estratégia de leitura no processamento de palavras. Sabendo-se que o processamento de palavras é fundamental no processo de leitura, cabe explorar o quanto práticas

pedagógicas voltadas para metacognição morfológica podem contribuir para a proficiência leitora.

O capítulo VII “O desenvolvimento da compreensão leitora: Cloze e atividades práticas no Ensino Fundamental”, de Kátia Abreu e Katharine da Hora, dá prosseguimento à investigação de natureza morfológica, mais especificamente, sobre a distinção entre itens funcionais e lexicais na língua. As autoras discutem a técnica de Cloze e apresentam as três fases do trabalho conduzido na escola, as quais, como nos demais estudos, envolveu uma fase inicial de diagnóstico, oficinas e retestagem. A técnica de Cloze é usada tanto com finalidade diagnóstica como interventiva, o que aproxima esse trabalho dos conduzidos com rastreador ocular. As autoras reportam resultados de estudo prévio que indicam a sensibilidade do teste de Cloze para avaliação de fatores linguísticos e não linguísticos na leitura (ABREU *et al.*, 2017) e apontam para o potencial do teste no acesso à proficiência leitora em níveis mais lexicais ou em níveis mais macroestruturais. Em resposta a críticas feitas ao emprego do teste de preenchimento de lacunas para avaliação de compreensão além do nível do período, Abreu e da Hora citam Gellert e Elbro (2013), os quais consideram que o trato na seleção das lacunas é que permitirá avaliar a compreensão das ideias no nível do período (local contexto) ou das ideias dos parágrafos. O teste diagnóstico foi conduzido com alunos do 8º ano do EF do CEJOTA e o texto que serviu de base para o Cloze foi uma fábula curta, extraído de livro didático de 8º ano. Tipo de palavra – lexical vs. funcional – foi definido como variável independente, e número de erros foi a variável dependente. Metade dos alunos preencheu lacunas com itens lexicais e a outra metade com itens funcionais. Embora a média de erros tenha sido maior para os itens funcionais, a diferença entre as condições não foi significativa. As oficinas realizadas envolveram tarefas de ordenação de palavras e expressões para composição de períodos, e também manipulação de conectivos. A retestagem foi realizada com alunos que participaram das oficinas e com um grupo controle de alunos da escola, de mesmo ano, que foram expostos à metodologia tradicional de ensino. Um texto similar ao empregado no teste diagnóstico foi utilizado. O experimento teve o mesmo *design* e os mesmos procedimentos foram adotados. Segundo as autoras, os resultados indicam que os alunos das oficinas tiveram sua compreensão leitora aprimorada e que a variação na acurácia das respostas desse grupo foi menor do que a do grupo controle. O caráter lúdico e interativo das oficinas e o prazer despertado pelas atividades realizadas são também apontados como fator importante na melhora de desempenho dos participantes.

O livro se encerra com o capítulo “Avaliação da competência leitora no ensino fundamental através da leitura auto-monitorada”, de Cristiane Ramos de Souza, Aleria Lage e Aniela Improta França. As autoras promovem uma reflexão sobre a realidade educacional brasileira, com foco na questão do alfabetismo, e trazem um dado muito preocupante – o de que 70% dos casos de analfabetismo no

país referem-se a crianças que frequentam a escola, mas que não se alfabetizam. O que justifica esse quadro e que caminhos podem ser estabelecidos para mitigá-lo? O trabalho das pesquisadoras se soma aqui ao dos apresentados nos demais capítulos do livro e representa uma tentativa de buscar resposta a essas perguntas. Conforme hipótese do neurocientista Dehaene e colaboradores (COHEN; DEHAENE, 2004; DEHAENE, 2009), durante a alfabetização, ocorre um processo de reciclagem neuronal - parte dos neurônios do córtex occipito-temporal ventral, antes voltados à interpretação de faces e objetos, passam a se especializar em decifrar as “linhas” dos grafemas, e a região onde ocorre essa reciclagem recebe o nome de Área da Forma Visual da Palavra (*Visual Word Form Area – VWFA*, em inglês). Segundo Ramos, Lage e França, a escola tem falhado em permitir que esse processo se dê de forma adequada. A preocupação com essa situação é a força motriz que guia o trabalho das autoras. Estas reveem estudos neurofisiológicos, de leitura automonitorada e de Cloze que levaram a efeito em uma fase piloto do projeto desenvolvido pelo laboratório LER e detêm-se, no capítulo, no relato do trabalho envolvendo o teste de leitura automonitorada⁵, entendido pelas pesquisadoras como uma ferramenta simples e eficiente, ao alcance de professores e alunos. No teste, os estímulos lidos pelos alunos foram sentenças simples do tipo sujeito-verbo-objeto, com um adjunto no início da frase e ao final. Esses mesmos estímulos foram usados em experimento anterior com EEG-ERP. As autoras desejavam comparar as duas técnicas, de modo a poder avaliar o poder de avaliação de cada uma. Aspectos semânticos (congruência entre o verbo e seu DP complemento – *No ônibus, o jogador comeu a maçã sem pressa vs. No ônibus, o jogador comeu a pedra sem pressa*) e aspectos associados ao visual externo do substantivo no DP complemento, chamada condição mancha (*maca* no lugar de *maçã*) ou ordenação dos grafemas da palavra do DP complemento (*mçaã* no lugar de *maçã*) foram manipulados. O experimento foi realizado com alunos do 8º ano do EF do CEJOTA. Os resultados da comparação das duas técnicas indicaram que a leitura automonitorada apresentou poder de avaliação similar ao do EEG-ERP, um teste muito mais custoso de ser aplicado em sala de aula. A comparação dos resultados da testagem e da retestagem mostrou que as oficinas realizadas no âmbito do projeto de leitura do LER impactaram o desempenho dos alunos – na retestagem, foi verificado efeito principal para tipo de DP e houve uma redução importante no tempo associado às leituras das condições congruente e incongruente. Além disso, o contraste entre a condição congruente e as duas condições incongruentes (uma delas com similaridade ortográfica – a condição mancha) aponta para acurácia leitora. A tarefa *off-line* associada aos dois testes – dizer se as sentenças faziam sentido (com as opções sim, não e talvez) -, também mostrou efeito das oficinas e destaca-se, na retestagem, o

5 Nesse tipo de teste, sentenças são lidas na tela de um computador, incrementalmente, em segmentos (em geral, palavra a palavra). O leitor controla a nova palavra a ser lida utilizando, para isso, a barra de espaço do computador. O tempo que o leitor leva para apertar a barra de espaço pode ser indicativo de custo cognitivo associado ao processamento da sentença em cada ponto da leitura.

uso da opção talvez, praticamente não escolhida na fase de diagnóstico, o que pode ser indicativo de início de metacognição. Segundo as autoras, o percentual de alunos identificados como bons leitores no teste diagnóstico foi de 28% e passou para 48% ao final das oficinas, na retestagem. Tem-se, pois, neste último capítulo, mais uma evidência de como técnicas e recursos usados em laboratórios psicolinguísticos podem ser levados para a sala de aula e usados de forma efetiva.

À guisa de conclusão, podemos dizer que a obra “Psicolinguística e Metacognição na Escola” é ela própria um instrumento, uma ferramenta de metacognição que leva o leitor a examinar o seu próprio modo de pensar o ensino de língua na escola. Ao trazer a perspectiva laboratorial para a sala de aula e propor a adoção de metodologias que implicam a participação ativa e reflexiva dos alunos, busca apontar caminhos que possam promover o acionamento do raciocínio científico e do pensamento crítico desses cidadãos em formação.

REFERÊNCIAS

ABREU, Kátia Nazareth Moura de Abreu; CID, Daniela de Garcia; HORA, Katharine da; SOUZA, Cristiane Ramos de. O teste de Cloze como instrumento de medida da proficiência em leitura: fatores linguísticos e não linguísticos. *Revista de Estudos da Linguagem*, v. 25, n. 3, p. 1767-1799, 2017. Disponível em <http://periodicos.letras.ufmg.br/index.php/relin/article/view/11366>. Acesso em abril de 2020.

ASTINGTON, Janet Wilde; GOPNIK, Alison. Theoretical explanations of children’s understanding of mind. *British Journal of Developmental Psychology*, v. 9, p. 7-31, 1991.

COHEN, Laurent; DEHAENE, Stanislas. Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, v. 22, p. 466-476, 2004.

DEHAENE, Stanislas. *Reading in the brain: the science and evolution of a human invention*. New York, Viking, 2009.

FLAVELL, John H. Metacognitive aspects of problem solving. In: RESNICK, Lauren B. (Ed.). *The nature of intelligence*. Hillsdale, NJ: Erlbaum, 1976, p. 231-235.

FRAZIER, Lyn; FODOR, Janet Dean. The Sausage Machine: A new two-stage parsing model. *Cognition*, v.6, p. 291-326, 1978.

GELLERT, Anna. S.; ELBRO, Carsten. Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational*

Assessment, v. 31, n. 1, p. 16–28, 2013.

HONDA, Maya; O'NEIL, Wayne. Triggering Science-Forming Capacity through Linguistic Inquiry. In: HALE, Kenneth; KEYSER, Samuel Jay (Eds.). *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, 1993, p. 229–255.

_____. *Thinking linguistically: A scientific approach to language*. Malden, MA: Blackwell Publishers, 2008.

_____. On thinking linguistically. *Revista Linguística*, Revista do Programa de Pós-Graduação em Linguística da Universidade Federal do Rio de Janeiro, v. 13, n.1, p. 52-65, 2017.

MAIA, Marcus Antonio. *Manual de linguística: subsídios para formação de professores indígenas na área de linguagem*. Brasília: UNESCO/MEC, 2006.

_____. *Um microscópio para ajudar a ler e escrever? Rastreamento ocular na leitura de alunos da educação básica*. Rede Nacional de Ciência para a Educação, 2017. Disponível em: <http://cienciaparaeducacao.org/blog/2017/03/13/conecta-eye-tracker-um-microscopio-para-ajudar-a-ler-e-escrever/>. Acesso em: maio de 2020.

_____. Computação estrutural de conjunto na leitura de períodos: um estudo de rastreamento ocular. In: MAIA, M. (org.). *Psicolinguística e educação*. Campinas: Mercado de Letras, 2018, p. 103-132.

MASON, Lucia; PLUSHINO, Patrik; TORNATORA, Maria Catarina. Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology*, v. 47, n. 6, p. 1083-1095, 2016.

OECD. *PISA 2018 Results (Volume I): What Students Know and Can Do*, OECD Publishing, Paris, 2019. Disponível em: <https://www.oecd-ilibrary.org/docserver/5f07c754-en.pdf?expires=1587837169&id=id&acname=guest&checksum=9E74C3B82CD8A4F27349706DCDD27155>. Acesso em: fevereiro de 2020.

PILATI, Eloisa *Linguística, gramática e aprendizagem ativa*. Campinas: Pontes, 2017.

RIBEIRO, Sara Bezerra dos Santos *Processamento da oração principal em períodos compostos por subordinação e coordenação: padrões de leitura e formação de pontos de vista*. Dissertação de Mestrado. Rio de Janeiro: Dep. de Pós-Graduação em Linguística, Universidade Federal do Rio de Janeiro, 2017.

SAVIANI, Dermeval O choque teórico da Politecnia. *Trabalho, Educação e Saúde*, v.1, n.1, p.131-152, 2003.

SILVA, Maria Cristina Figueiredo; MEDEIROS, Alessandro Boechat de. *Para conhecer Morfologia*. São Paulo: Contexto, 2016.

SILVANO, Emily da Silva. *Rastreamento ocular de orações temporariamente ambíguas: reflexões sobre os efeitos garden-path e good-enough na leitura de alunos de curso fundamental e superior*. Dissertação de Mestrado. Rio de Janeiro: Dep. de Pós-Graduação em Linguística, Universidade Federal do Rio de Janeiro, 2018.