

Universidade Federal de Juiz de Fora  
Programa de Pós-Graduação em Linguística  
Doutorado em Linguística

Milene Cristine de Castro Teixeira Laguardia

ETAPAS INICIAIS DE AQUISIÇÃO LEXICAL – HABILIDADES ESTATÍSTICAS E  
SIMBÓLICAS NO TRATAMENTO DE DEPENDÊNCIAS NÃO ADJACENTES

Juiz de Fora

2016

MILENE CRISTINE DE CASTRO TEIXEIRA LAGUARDIA

**ETAPAS INICIAIS DE AQUISIÇÃO LEXICAL – HABILIDADES ESTATÍSTICAS E  
SIMBÓLICAS NO TRATAMENTO DE DEPENDÊNCIAS NÃO ADJACENTES**

Tese apresentada ao Programa de Pós-Graduação em Linguística da UFJF, área de concentração: Linguística e Cognição, como requisito parcial para obtenção do título de Doutor em Linguística.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Cristina Lobo Name

Juiz de Fora

Outubro de 2016

Ficha catalográfica

Folha de aprovação

## RESUMO

Esta tese tem como objetivo geral investigar as etapas iniciais da aquisição lexical, focalizando a capacidade de o bebê identificar, abstrair e generalizar padrões de dependências não-adjacentes (DNAs) nos enunciados linguísticos e de usá-los na categorização inicial de elementos do léxico. Nossa tese é que o estabelecimento de relações entre elementos de DNAs é parte da etapa inicial de aquisição lexical e que, para tal estabelecimento, além de habilidades perceptuais, faz-se necessário o uso de habilidades computacionais – de manipulação de símbolos através de algoritmos – que torne possíveis (i) o reconhecimento e a identificação de elementos inter-relacionados, que se apresentam em configuração de DNAs, e (ii) a abstração e generalização desses padrões a novos estímulos. DNAs são definidas como a co-ocorrência de marcadores morfossintáticos e/ou fonológicos, com material interveniente. Elegeram-se as DNAs como objeto de estudo por serem altamente frequentes nas línguas naturais e expressarem relações morfossintáticas (p.ex., concordância verbal, nominal de gênero e/ou de número etc.). Trabalhos com línguas artificiais (SAFFRAN et al 96; GOMEZ et al. 2006; GOMEZ & MAYE, 2005; GOMEZ & LAKUSTA, 2004; LANGUS et al 2012 ; VAN DEN BOS et al 2012) sustentam que habilidades estatísticas e probabilísticas, de domínio geral, subjazem ao tratamento de DNAs por bebês e adultos. Nossa hipótese, em consonância com Marcus et al. (1999) e Peña et al. (2002), é que tais habilidades, sozinhas, não dão conta do tratamento de DNAs em língua natural e, portanto, bebês recorrem tanto a mecanismos gerais quanto a mecanismos computacionais/simbólicos, especificamente linguísticos, na aquisição de uma língua. Duas atividades experimentais foram desenvolvidas, com os seguintes objetivos específicos: (i) investigar se em ambiente prosódico determinado – alinhadas à fronteiras de sintagmas fonológicos – DNAs seriam mais facilmente percebidas por bebês do que em ambientes em que estão distantes de fronteiras fonológicas (Experimento 1 – participantes: bebês canadenses e bebês brasileiros, com média de idade de 11 meses); (ii) investigar se bebês são capazes de categorizar linguisticamente os elementos do léxico a partir do reconhecimento, abstração e generalização de DNAs (Experimento 2- participantes: bebês brasileiros, com média de idade de 11 meses). Nossos resultados são discutidos a partir de uma abordagem mista/integrada, que assume habilidades gerais e especificamente linguísticas no tratamento de enunciados pelo bebê durante o processo inicial de aquisição (GERVAIN & MEHLER, 2010; ENDRESS, NESPOR & MEHLER, 2009) e na hipótese do Bootstrapping prosódico (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997, 2008) e indicam que um ambiente prosódico específico (presença de fronteiras de sintagmas fonológicos) favorece o reconhecimento, abstração e generalização de DNAs e que estas favorecem a categorização inicial de itens do léxico.

**Palavras-chave:** Aquisição lexical; dependências não-adjacentes; categorização; habilidades estatísticas e computacionais; Psicolinguística.

## ABSTRACT

This thesis is concerned with the investigation of initial stages of language acquisition by focusing on infant's ability to identify, abstract and generalize non adjacent dependencies (NAD) patterns in the linguistic stream and also use them in the initial categorization of lexical elements. Our thesis is that the establishment of relations between NAD is part of initial lexical acquisition and for this establishment infants use perceptual and computational abilities – to treat symbols beyond algorisms – and that comes possible: (i) the recognition and identification of related terms which appears in NAD configuration, and (ii) the abstraction and generalization of these patterns to new stimulus. NAD are the co-occurrence of morphosyntactical and/or phonological markers with another element between them. NAD was chosen as study object because they are highly frequent elements in natural languages and they express morphosyntactic relations (like verbal, noun or gender agreement). Studies about artificial languages (SAFFRAN et al 96; GOMEZ et al. 2006; GOMEZ & MAYE, 2005; GOMEZ & LAKUSTA, 2004; LANGUS et al 2012 ; VAN DEN BOS et al 2012) assume that statistical and probabilistic abilities, which belong to general domain, underlie the NAD treatment by infants and adults. Our hypothesis, according Marcus et al. (1999) e Peña et al. (2002), is that this kind of ability is not enough to the treatment of NAD in natural language and so infants use as domain general mechanisms as well computational/symbolic, domain especific, to acquire a language. Therefore, two experiments are reported with the following specific purposes: (i) to explore if a determined prosodic context – alligned to phonological phrase boundary – NAD is easier perceived by infants than in context which NAD is far to phonological phrase boundary (Experiment 1 – canadian infants and brazilian infants, medium age of 11 months). (ii) to explore if infants are able to categorize (linguistic categorization) lexical elements through the NAD recognition, abstraction and generalization (Experiment 2 – brazilian babies, medium age 11 months). Our results are discussed based on a mixed/ integrated approach with assumes domain general abilities and specific abilities in the processing of statements by infants during the initial language process (GERVAIN & MEHLER, 2010; ENDRESS, NESPOR & MEHLER, 2009) and on the prosodic Bootstrapping hypothesis (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997, 2008) and suggest that a specific prosodic context ( prosodic phrase boundary alligned to NAD) promotes NAD recognize, abstraction and generalization and suggest that theses ones promotes the initial categorization of lexical itens.

**KEYWORDS:** Lexical acquisition; non adjacent dependencies; categorization; statistic and computational abilities; Psycholinguistics

## **LISTA DE FIGURAS**

Figura 1 – Modelo dos estágios iniciais do processamento de fala (Christophe <i>et al.</i> , 2008).....	37
Figura 2 - Esquema extraído de Endress et al. 2009, pág. 349.....	50
Figura 3 – Espectograma do estímulo “Essa dabo munil” na condição Adjetivo .....	62
Figura 4 - Espectograma do estímulo “Essa dabo muniu” na condição Verbo .....	63
Figura 5 – Imagem Familiarização e Teste .....	72
Figura 6 – Imagem “attention getter”.....	72
Figura 7 - Cabine de atividades experimentais onde as atividades foram realizadas na UFJF. ....	73
Figura 8 - Cabine de atividades experimentais onde as atividades foram realizadas na UQAM.....	73

## **LISTA DE GRÁFICOS**

Gráfico 1 – Alongamento das fronteiras de sintagma fonológico nas Condições Verbo e Adjetivo ....	66
Gráfico 2 – Resultado com bebês canadenses- Tempo médio em cada condição.....	75
Gráfico 3 – Resultado com bebês brasileiros- Tempo médio em cada condição.....	82
Gráfico 4 – Resultado média tempo de escuta experimento categorização .....	90

## **LISTA DE TABELAS**

Tabela 1 - Combinação entre Determinantes e terminação dos N na condição Verbo .....	60
Tabela 2 - Combinação entre Determinantes e terminação dos N na condição Adjetivo .....	61
Tabela 3 – Média dos valores de Duração, Intensidade, Intensidade Mínima e Intensidade Máxima da última sílaba do Nome .....	64
Tabela 4 – Média dos valores de Duração, Intensidade, Intensidade Mínima e Intensidade Máxima da primeira sílaba do Adjetivo/Verbo.....	65
Tabela 5 – Visualização das combinações entre Det e terminação de N nas fases de familiarização...	71
Tabela 6 – Teste de normalidade das amostras das condições Adjetivo e Verbo – Experimento 1A ...	76
Tabela 7 - Retas de normalidade das condições Adjetivo e Verbo- Experimento 1A .....	76
Tabela 8 -Teste de normalidade das amostras das condições Adjetivo e Verbo – Experimento 1B....	80
Tabela 9 - Retas de normalidade das condições Adjetivo e Verbo- Experimento 1B.....	81
Tabela 10 - Esquema das combinações utilizadas na atividade .....	86
Tabela 11– Teste de normalidade das amostras do Experimento de categorização .....	89
Tabela 12- Retas de normalidade experimento Categorização .....	89

## **LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS**

Adj Adjetivo

D / Det Determinante

DA Dependências adjacentes

DNA Dependências não adjacentes

DP Determinant Phrase

ERPs Event-related brain potentials

f0 Frequência fundamental

N Nome

NP Nominal Phrase

PB Português brasileiro

PM Programa Minimalista

POMPs *Perceptual or memory primitives*

V Verbo

VP Verbal Phrase

∅ Sintagma fonológico

## SUMÁRIO

INTRODUÇÃO .....	12
1.1- Objetivos .....	16
1.2- Organização do trabalho.....	17
2- DEPENDÊNCIAS NÃO ADJACENTES .....	18
2.1- Caracterização.....	18
2.1.1- Evidências de sensibilidade às DNAs.....	22
2.2- Elementos funcionais .....	25
2.2.1- Evidências de sensibilidade aos elementos funcionais .....	27
2.3- Pareamento DNAs e constituintes prosódicos .....	29
2.3.1 - O sintagma fonológico ( $\phi$ ) .....	32
2.4- Conclusão.....	34
3- HABILIDADES DE SEGMENTAÇÃO E IDENTIFICAÇÃO DE PADRÕES .....	36
3.1- O Bootstrapping prosódico .....	36
3.2- Habilidades de identificação de padrões.....	38
3.3- Habilidades de domínio geral x Habilidades de domínio específico .....	43
3.4- Proposta de integração .....	48
4- ATIVIDADES EXPERIMENTAIS .....	53
4.1- Técnica Experimental – Olhar preferencial .....	54
4.2- Atividade Experimental 1 - Relação entre DNA e prosódia (1A) .....	57
4.2.1- Elaboração dos estímulos.....	59
4.2.2 – Análise acústica dos estímulos .....	61
4.2.3- Variáveis experimentais:.....	66
4.2.4- Método .....	67
4.2.5 -Resultados e Discussões (1A) .....	74
4.2.6- Experimento prosódia bebês brasileiros (1B) .....	78
4.3- Atividade experimental 2 – Capacidade de categorização.....	84
4.3.1- Elaboração dos estímulos.....	85
4.3.2- Variáveis experimentais.....	87
4.3.3- Resultados e Discussões .....	88
4.4 – Como avaliar os dados perceptuais do bebê durante a atividade experimental.....	92
5- CONCLUSÃO .....	95
6- REFERÊNCIAS.....	102

Glossário .....	108
Anexos.....	113

## INTRODUÇÃO

Esta tese focaliza as etapas iniciais de aquisição lexical, investigando a abstração e generalização de dependências não-adjacentes (doravante, DNAs) e a categorização de itens da língua a partir da integração de mecanismos estatísticos e de habilidades computacionais/simbólicas para o tratamento das informações de natureza linguística. Dependências não adjacentes são definidas como a co-ocorrência de marcadores morfossintáticos e/ou fonológicos com material interveniente e são constituídas, em sua grande maioria, de itens funcionais, como em “**Ela fala**” e **Nós dançamos**”. Como veremos, têm papel crucial na aquisição lexical e sintática, à medida que manifestam relações sintáticas (concordância) e permitem a identificação de categorias gramaticais. Por abstração e generalização<sup>1</sup> entende-se que são habilidades de reconhecimento de novas sequências como exemplares de um padrão aprendido. Ainda, categorização “é o processo pelo qual entidades distintas são tratadas como equivalentes e é uma das atividades cognitivas mais fundamentais, pois nos permite compreender e fazer previsões sobre objetos e eventos do mundo” (MEDIN; AGUILAR, 1999, p. 104).<sup>2</sup>

Nossa tese é que o estabelecimento de relações entre elementos de DNAs é parte da etapa inicial de aquisição lexical e que, para tal estabelecimento, além de habilidades perceptuais, faz-se necessário o uso de habilidades computacionais – de manipulação de símbolos através de algoritmos – que torne possíveis (i) o reconhecimento e a identificação de elementos inter-relacionados, que se apresentam em configuração de DNAs, e (ii) a abstração e generalização desses padrões a novos estímulos. Como veremos, DNAs podem ser percebidas através de mecanismos estatísticos, mas assumimos que a capacidade de abstrair padrões de DNAs e generalizá-los depende de capacidades computacionais e não meramente estatísticas. Na literatura, como será apresentado ao longo da tese, as habilidades estatísticas têm sido propostas como habilidades gerais – não específicas para a linguagem, e o

<sup>1</sup> Não há na literatura definições bem estabelecidas desses termos. Além disso, de acordo com Jenny Saffran, em conversas trocadas por e-mail, os dois termos têm sido usados de modo intercambiável. Neste trabalho, entretanto, optamos por usar os dois termos por acreditar que, ainda que intercambiáveis na literatura, a generalização de um padrão parece exigir que a abstração do mesmo tenha ocorrido.

<sup>2</sup> Para um melhor entendimento do que estamos assumindo, os principais conceitos utilizados nesta tese - **aquisição lexical, abstração, generalização, categorização, aprendizagem estatística e habilidades computacionais/simbólicas** – estão definidos no Glossário.

tratamento simbólico dos estímulos linguísticos tem sido proposto como de domínio específico. Esta tese defende, portanto, que na aquisição de uma língua, bebês recorram tanto a mecanismos não especificamente linguísticos, pelo menos inicialmente, quanto a mecanismos especificamente linguísticos.

No âmbito da teoria linguística, a vertente gerativista (Chomsky 1965 e obras posteriores), comprometida desde sua origem com o problema da aquisição da linguagem em uma perspectiva que trata a capacidade linguística como um sistema cognitivo biologicamente especificado, será assumida como referência nesta tese. Assumimos ainda o que chamamos de uma abordagem mista/integrada ao adotarmos a proposta do *Bootstrapping* Prosódico (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997) e uma proposta de integração mista no tratamento do estímulo linguístico (GERVAIN & MEHLER, 2010) para o processo inicial de aquisição lexical. De acordo com a hipótese do *Bootstrapping* Prosódico, a análise fonológica do sinal da fala e a existência de pistas variadas, como pistas distribucionais, pistas de fronteiras prosódicas, regularidades fonotáticas, ocorrência de marcadores morfossintáticos em ambientes frequentes, podem permitir às crianças adquirirem o léxico e a sintaxe de sua língua. Por sua vez, diferentemente de alguns autores, não estamos defendendo que uma perspectiva exclusivamente estatística para a linguagem dê conta do processo de aquisição lexical. Assim como propõem Gervain e Mehler (2010), consideramos a necessidade de múltiplos fatores como, por exemplo, a capacidade inata de adquirir uma língua natural, habilidades estatísticas e o uso de habilidades computacionais específicas para a linguagem – tratamento simbólico – de igual maneira necessárias para a aquisição da língua.

A partir dessa abordagem, buscamos, além de estudar as capacidades de abstração e generalização de padrões linguísticos por bebês durante o primeiro ano de vida, investigar os fatores facilitadores desse processo, assim como o uso de tais capacidades na categorização inicial de elementos do léxico. Como veremos, há fatores/ambientes linguísticos – como, por exemplo, um ambiente prosódico específico (atividade experimental 1) –, que favorecem o mapeamento e abstração de DNAs. Ainda, mapeamento, abstração e generalização contribuem para a categorização dos elementos lexicais, mesmo que essa categorização seja “inicial” e estritamente (morfo)sintática, i.e., não conceitual/semântica (atividade experimental 2).

De maneira geral, como também veremos adiante, diversos autores têm explorado as capacidades de abstração e generalização de padrões linguísticos tanto por adultos quanto por crianças. Entretanto, os que trabalham com línguas naturais, investigam tais capacidades em crianças no seu segundo ano de vida (18 meses), considerando sobretudo o reconhecimento de

padrões de sua língua materna, e não propriamente habilidades de abstração e generalização (SANTELmann & JUSCZYK, 1998; HOHLE & WEISSENBORN, 2006; VAN HEUGTEN & SHI, 2010). Os que utilizam línguas totalmente artificiais (MARCUS et al., 1999; GÓMEZ, 2002; NEWPORT & ASLIN, 2004) não consideram propriedades morfofonológicas e prosódicas intrínsecas aos enunciados disponíveis à criança. Assim, nos primeiros estudos temos evidência de reconhecimento de sua língua materna pelo bebê, mas não de tratamento simbólico; por outro lado, os resultados do outro grupo de trabalhos apontam tratamento simbólico dos estímulos, mas não podem ser generalizados para língua natural, visto que os estímulos não apresentam as propriedades inerentes às línguas naturais.

Para defender que o estabelecimento de relações entre itens de DNAs faz parte do processo de aquisição, precisamos investigar a percepção, generalização e abstração de DNAs em uma língua com configuração próxima a uma língua natural, mantendo, ao menos, características que esta apresenta. Isso é o que esta tese propõe. Vale ressaltar que a língua utilizada em nossas atividades não é totalmente natural, pois há diminuição entre as combinações possíveis e existentes em uma língua natural, mas por outro lado, foram mantidas propriedades inerentes às línguas naturais, como distinções prosódicas e fonotáticas dos itens funcionais (presentes nas DNAs) e lexicais e o contorno entoacional característico dos constituintes – diferentemente das línguas totalmente artificiais. Dessa forma, a pseudolíngua utilizada, criada a partir do Português do Brasil (PB), é uma tentativa de apresentar aos bebês uma língua com características mais próximas, dentro do possível, das línguas naturais, mas ainda de manter um ambiente propício para a investigação das habilidades de tratamento simbólico, o que ainda não foi apresentado na literatura.

Name, Shi e Koulaguina (2011) sugerem que bebês canadenses de 11 meses, expostos durante um curto período de tempo (2 minutos) a uma língua pseudonatural semelhante ao português brasileiro (PB) – portanto, diferente de sua língua materna –, são capazes de reconhecer padrões de relações entre determinantes e nomes e, ainda, são capazes de abstraí-los de modo a identificá-los em novos enunciados. A mesma atividade foi desenvolvida com bebês brasileiros no mestrado da autora desta tese (TEIXEIRA, 2012; NAME & LAGUARDIA, 2013); entretanto, consideramos que para esse grupo de participantes a tarefa foi mais complexa, visto que os bebês tiveram que abstrair uma nova subclassificação para determinantes já conhecidos, devido à exposição prévia ao PB. Os resultados sugerem que aos 11 meses, mesmo em uma tarefa mais difícil em relação aos bebês canadenses, os bebês foram eficientes na tarefa de reconhecer padrões de relações entre D e N, abstraí-los e generalizá-los em novos enunciados nunca antes ouvidos. A partir desse estudo, vários

questionamentos surgiram, nos desafiando a ampliar essa pesquisa. Algumas indagações sobre qual poderia ser a influência de um ambiente prosódico específico nesta tarefa e sobre a possibilidade das crianças categorizarem esses elementos abstraídos foram motivadoras a fim de propor um conjunto de pistas e de informações das quais as crianças poderiam usufruir nas etapas iniciais da aquisição lexical.

A partir dessas indagações, buscamos abordar o processo inicial de aquisição de uma língua pela criança, especificamente a abstração e generalização de dependências não-adjacentes entre D e terminação de N e a categorização de itens da língua, considerando as habilidades que se postulam necessárias para que a criança possa identificar a língua de seu meio e as propriedades que essa língua apresenta, para que seja possível de ser adquirida espontaneamente.

A percepção de características acústicas dos constituintes prosódicos parece facilitar a segmentação do enunciado em unidades menores pelo bebê. Pesquisas experimentais apresentam evidência de sensibilidade a propriedades prosódicas do estímulo linguístico que são usadas precocemente pelo bebê nas etapas iniciais da aquisição lexical. Além disso, o reconhecimento de itens funcionais pode ajudar os bebês na segmentação e categorização dos itens lexicais (JUSCZYK, CUTLER & REDANZ, 1993; TURK, JUSCZYK & GERKEN, 1995). No que tange às DNAs, a identificação dos itens funcionais torna-se primordial para o reconhecimento, abstração e generalização dos padrões de DNAs. Uma série de trabalhos aponta para sensibilidade precoce aos itens funcionais e à sua posição por bebês adquirindo línguas variadas (SHADY, 1996; SHAFFER et al., 1998, entre outros).

O desenvolvimento deste trabalho justifica-se pela intenção de se ampliar a pesquisa de Teixeira (2012) em que se investigou a contribuição das informações distribucionais de DNA nas etapas iniciais da aquisição lexical. O intuito do presente estudo é ampliar esses resultados, investigando i) se um contorno prosódico específico poderia “facilitar” a abstração e generalização de DNAs; ii) se a abstração e generalização de DNAs podem facilitar a categorização de itens do léxico e iii) como a integração entre mecanismos estatísticos e habilidades computacionais podem contribuir para o tratamento do estímulo linguístico. O presente estudo se insere no projeto “Relações de coocorrência em constituintes prosódicos na aquisição lexical”, desenvolvido no Núcleo de Estudos em Aquisição da Linguagem e Psicolinguística – NEALP – da Universidade Federal de Juiz de Fora.

### **1.1- Objetivos**

O objetivo geral dessa tese é investigar as etapas iniciais da aquisição lexical, focalizando a abstração e generalização de dependências não-adjacentes e a categorização de itens das categorias Determinante (D) e Nome (N), a partir da integração de tratamentos estatísticos e de habilidades computacionais/simbólicas para o tratamento das informações linguísticas.

Seus objetivos específicos são:

- 1- Investigar se um ambiente prosódico específico pode facilitar o aprendizado de DNAs. Em consonância com esse objetivo, uma atividade experimental foi proposta (Experimento 1) com o intuito de investigar se DNAs alinhadas a fronteiras de sintagmas fonológicos poderiam ser mais facilmente percebidas por bebês aos 11 meses do que em ambientes em que estão distantes de fronteiras fonológicas.
- 2- Investigar se bebês são capazes de categorizar linguisticamente elementos do léxico a partir do reconhecimento, abstração e generalização de DNAs (Experimento 2).
- 3- Discutir em que medida a proposta de integração entre habilidades estatísticas e computacionais/simbólicas surge como uma proposta possível para o tratamento das informações linguísticas na aquisição de uma língua natural.
- 4- Articular os resultados encontrados nas atividades experimentais com a proposta de integração mista assumida nesta pesquisa, discutindo em que medida a integração entre habilidades de tratamento estatístico e habilidades computacionais/simbólicas são determinantes no processo inicial de aquisição lexical e, mais especificamente, no mapeamento, abstração e generalização de DNAs.

## 1.2- Organização do trabalho

Os capítulos desta tese serão divididas da seguinte maneira:

O segundo capítulo será dedicado à caracterização das dependências não adjacentes – objeto de nosso estudo, sua importância nas línguas naturais e no processo inicial de aquisição lexical e ainda evidências de sensibilidade às DNAs. Em continuidade, serão apresentadas as características dos itens funcionais e evidências de sensibilidade aos mesmos, visto que DNAs são constituídas, majoritariamente, de elementos de itens funcionais e, por último, o pareamento entre DNAs e constituintes prosódicos, destacando o sintagma fonológico, destacando que esse foi o ambiente prosódico ao qual DNAs foram alinhadas em nossa atividade experimental 1.

O terceiro capítulo se dedicará às habilidades de segmentação e identificação de padrões, à apresentação do modelo de *Bootstrapping Prosódico* (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997) e à dualidade entre habilidades de domínio geral e habilidades de domínio específico. Neste capítulo também será apresentada a proposta de integração mista para tratamento do estímulo linguístico (GERVAIN & MEHLER , 2010) para o processo inicial de aquisição lexical, assumida nesta tese.

O quarto capítulo apresentará a metodologia e as atividades experimentais desenvolvidas (Experimentos 1 e 2), bem como a discussão dos resultados obtidos.

O quinto capítulo será dedicado à uma síntese da tese, conciliando a hipótese de trabalho, os objetivos propostos e os resultados obtidos e buscando propor em que medida eles podem ser discutidos à luz da proposta de integração assumida e em que medida podem contribuir para a mesma.

## 2- DEPENDÊNCIAS NÃO ADJACENTES

Na introdução, vimos que o objetivo desta tese é investigar a abstração e generalização de DNAs e a categorização de itens da língua a partir da integração de mecanismos estatísticos e habilidades computacionais para o tratamento de informação linguística nas etapas iniciais de aquisição lexical.

Neste capítulo, apresentaremos a caracterização das DNAs ressaltando dois aspectos principais, os quais são centrais nesta tese: i) DNAs são constituídas por elementos funcionais (itens funcionais, *word markers* e base lexical interveniente) e ii) DNAs se apresentam frequentemente em fronteiras de domínios prosódicos, como o grupo clítico e o sintagma fonológico. Assim, após a caracterização das DNAs e apresentação das evidências de sensibilidade a esse fenômeno, será apresentada uma caracterização dos elementos funcionais, destacando sua importância para o processo de aquisição de uma língua e apresentando evidências de sensibilidade aos mesmos. Logo em seguida, será apresentada uma caracterização do pareamento entre DNAs e constituintes prosódicos, seguida de evidências de sensibilidade a esse tipo de pareamento.

### 2.1- Caracterização

As dependências não adjacentes são constituídas, em sua grande maioria, de itens funcionais e são definidas como a co-ocorrência de marcadores morfossintáticos e/ou fonológicos com material interveniente. A relação de coocorrência pode ocorrer entre itens funcionais, como em

- 1) “**Eu danço**”  
 Det Terminação verbal  
 (1<sup>a</sup>p. sing.) (1<sup>a</sup>p. sing.)

mas também entre um item funcional e um não funcional, como nos casos das vogais temáticas em:

2) “**A menina**”

Det Vogal temática  
(item funcional)

Referem-se, portanto, majoritariamente, mas não exclusivamente, a elementos de categoria funcional que estão relacionados, porém com um componente extra entre um elemento funcional e outro (funcional ou não funcional, como vogal temática de N).

Nas línguas naturais, apesar de serem recorrentes nos enunciados, são variáveis, o que poderia dificultar sua identificação. Em português, por exemplo, podemos ter diferentes formas de determinantes – o/um/esse/meu/nosso... – e de nomes – bolo/dente/sol/nariz... e mesmo que as crianças não tenham sido expostas a todas as combinações possíveis entre os itens das categorias determinantes e nomes, elas teriam de perceber as possíveis combinações entre os itens dessas categorias (NAME e LAGUARDIA, 2013).

As dependências não-adjacentes são consideradas de fundamental importância para o processo de aquisição de uma língua, pois:

- i) É um fenômeno presente em línguas naturais e manifesta relações importantes, como relações de concordância e de categorização e, devido a sua recorrência, é de fácil identificação no estímulo linguístico, facilitando a aquisição da linguagem.
- ii) Expressam a (sub)categorização de novas palavras, ainda que o material interveniente seja bastante variável, como o é em línguas naturais:

3) **Um X-o; X= N**

(Det masc.) (Nome com terminação  
tipicamente masc. )

Se X é precedido por Det e contém uma terminação compatível com o mesmo, então X é um Nome.

- iii) São auxiliadoras no processo de identificação de concordância de gênero e/ou número entre Determinante, Nome e Adjetivo; na concordância de sujeito e verbo, ou seja, encerram morfológicamente relações sintáticas de concordância, como nos exemplos abaixo

**4) Um livro**

(masc.sing) (masc.sing)

**5) As bolas coloridas**

(fem.plural) (fem.plural) (fem.plural)

**6) Eu canto**

(1<sup>a</sup>p. sing.) (1<sup>a</sup>p. sing.)

**7) Ele canta**

(3<sup>a</sup>p.sing) (3<sup>a</sup>p.sing)

**8) E ainda em : “... está correndo”**

(verbo “estar” presente 3<sup>a</sup>p. sing) (terminação gerúndio)

“ **tem corrido**”;

(verbo “ter” presente 3<sup>a</sup>p. sing) (terminação particípio)

Muitas são as línguas naturais que apresentam o fenômeno das DNAs para estabelecer relações morfossintáticas entre seus elementos. Na língua inglesa, por exemplo, dependências não adjacentes morfossintáticas podem ocorrer ao menos em dois casos:

(i) Dependência morfossintática de tempo, como a que foi utilizada no estudo de Santelmann & Jusczyk (1998) entre a forma verbal da terceira pessoa do presente do singular “*is*” (verbo *to be*) e terminação característica de gerúndio “*ing*”. Nota-se que quando um desses elementos é alterado (inserindo a forma “*can*”, por exemplo) a sentença se torna agramatical.

8) At the bakery, everybody **is baking** bread.

\*At the bakery, everybody **can baking** bread

(ii) Dependência morfossintática de concordância entre sujeito e verbo. Como, por exemplo, na concordância da terceira pessoa do singular e a terminação do verbo no presente.

- 9) **He** laughs (Ele gargalha)
- 10) **She** loves (Ela ama)
- 11) **They** run (Eles correm)

Os exemplos acima apresentam as DNAs em um mesmo constituinte fonológico. Entretanto, essa relação de co-dependência com elementos intervenientes também pode ocorrer em fronteiras de sintagmas fonológicos distintos, como nos exemplos em que o pronome demonstrativo “*these*” e o pronome definido “*the*” determinam a forma verbal do verbo “*to have*”

- 12) [**These** sheeps]  $\emptyset$  [**have** slept]
- 13) [**The** sheep]  $\emptyset$  [**has** slept].

Na língua francesa, assim como na língua portuguesa, a forma do pronome de 1<sup>a</sup> pessoa do plural determina a forma exata da flexão do verbo. Quando uma das duas formas é modificada, a sentença se torna agramatical (\*). Trata-se de uma relação de concordância entre qualquer verbo e seu sujeito.

- 14) **Nous** partons demain. (Nós partiremos amanhã)

\***Nous** partez demain.

\***Vous** partons demain.

Em italiano, há também uma relação de concordância entre o artigo e a terminação do nome. Assim, a flexão do nome determina o artigo que o acompanha.

- 15) **La** bambina (A menina)
- 16) **Il** bambino (O menino)

Em alemão, há a marcação dupla entre o domínio funcional de auxiliares (heb) e o domínio lexical do predicado verbal (ge) que é a marcação do particípio.

- 17) Vandaag **heb** ik de dokter **gebeld** (Hoje eu tive o médico chamado)

Como podemos observar nos exemplos citados, a relação de DNAs pode ocorrer entre itens de elementos funcionais, como entre o pronome e a marcação verbal (nas línguas inglesa e francesa), entre determinante e a flexão nominal (na língua italiana) e entre o auxiliar e o predicado verbal (na língua alemã). Os itens funcionais devido às suas características e possibilidade de se configurarem em relações de DNAs tornam-se elementos importantes para o processo de aquisição de uma língua natural. Na seção 2.2, apresentaremos, a caracterização dos elementos funcionais.

Vimos nesta seção uma caracterização das DNAs. Na próxima subseção apresentaremos evidências de sensibilidade à esse fenômeno.

### **2.1.1- Evidências de sensibilidade às DNAs**

Utilizando língua natural, Santelmann & Jusczyk (1998) focalizaram a sensibilidade de bebês americanos a DNAs entre o auxiliar *-is* e o morfema verbal *-ing* (na versão normal) e entre *can* e o morfema verbal *-ing* (versão modificada), utilizando a técnica de escuta preferencial. Crianças de 15 e de 18 meses escutaram pequenas histórias contendo verbos no presente contínuo (*is ...-ing*), na versão normal, e (*can ...-ing*) na versão modificada. Baseados em estudos anteriores (SHADY, 1996) em que crianças entre 14 e 16 meses mostraram sensibilidade à distribuição e função da ordem dos morfemas em inglês, nesse trabalho, bebês, aos 15 meses, não demonstraram sensibilidade a dependências entre o auxiliar “*is*” e o morfema “*ing*”, ou seja, não reconheceram a co-ocorrência entre esses morfemas. Entretanto, diferentemente do resultado obtido com bebês aos 15 meses, bebês aos 18 meses se mostraram sensíveis para as relações entre esses morfemas, estabelecendo a relação entre os elementos não contíguos presentes na língua inglesa. Os pesquisadores fizeram outros experimentos, variando o número de sílabas encaixadas entre os dois morfemas (colocando advérbios com duas ou mais sílabas), com crianças de 18 meses. As crianças continuaram sensíveis à relação entre os morfemas em uma janela de até 3 sílabas, i.e., quando havia entre *is* e *-ing* somente a sílaba do verbo ou, além dela, um advérbio de 2 sílabas. Porém, com uma “distância” maior (com advérbios de 3 e 4 sílabas), não houve diferença significativa no tempo médio de escuta das duas versões. Esses resultados apontam para a existência de uma “janela de processamento”, em que a criança é capaz de estabelecer relações entre morfemas.

Van Heugten & Shi (2010) com o objetivo de investigar se aprendizes são sensíveis a DNAs, quando estas extrapolam o domínio de fronteira fonológica, testaram bebês aprendizes da língua francesa, utilizando a técnica do Olhar preferencial (*Preferential Looking*). Quatro palavras alvo familiares com a inicial [ba-] foram selecionadas, duas carregando o gênero masculino (*ballon* ‘balão’, *bateau* ‘barco’) e duas carregando o gênero feminino (*banane* ‘banana’, *balançoire* ‘balanço’). As consoantes do onset na segunda sílaba dos alvos eram todas coronais. A escolha do alvo de ‘ba+coronal’ elimina a possibilidade dos efeitos de coarticulação do determinante precedente, o que poderia afetar nas respostas das crianças. Além dos nomes alvos, quatro outros nomes foram selecionados, dois masculinos (*crayon* ‘lápis’ e *livre* ‘livro’) e dois femininos (*table* ‘mesa’ e *fleur* ‘flor’). Os resultados indicam que aos 17 meses, mas não aos 14, DNAs são rastreadas mesmo quando o determinante e o verbo auxiliar pertencem a sintagmas fonológicos distintos, o que é notável, visto que os sintagmas fonológicos são consideradas as principais unidades de processamento para as crianças.

Em língua pseudonatural (i.e., criada, mas apresentando propriedades morfofonológicas e prosódicas presentes em língua natural) e trabalhando com crianças mais jovens, NAME, SHI & KOULAGUINA (2011), utilizando a técnica do Olhar Preferencial, sugerem que bebês canadenses de 11 meses, expostos durante um curto período de tempo (2 minutos) a uma língua pseudonatural – diferente de sua língua materna –, são capazes de reconhecer padrões de relações entre D e N e, ainda, são capazes de abstraí-los de modo a identificá-los em novos enunciados. Buscando dar continuidade aos resultados encontrados em estudo anterior da autora desta tese (TEIXEIRA, 2012) e utilizando a mesma técnica do experimento canadense, bebês brasileiros foram testados com média de idade de 11 meses, durante um curto período de tempo a DPs de uma língua pseudonatural (2 min), seguindo dois padrões – nosso/seu + N-e (ex.: “Nosso pebe, seu pebe”); esse/meu + N-a (ex.: “Esse tova, meu tova”). Eles foram expostos a novos DPs contendo os mesmos determinantes familiarizados seguidos de novos pseudónomes com terminação congruente às combinações apresentadas anteriormente (i.e., grammatical) ou incongruente (i.e., agrammatical). Os resultados obtidos a partir da média de tempo de escuta entre as condições gramaticais e agramaticais indicaram que os bebês reagiram à diferença de padrões de DNA em função daqueles previamente apresentados, escutando por mais tempo padrões não familiarizados, sugerindo que a partir de uma curta exposição a uma língua pseudonatural – diferente do padrão do PB – bebês, aos 11 meses, são capazes de abstrair e generalizar o padrão dessa “língua”. Tais resultados sugerem que mecanismos de abstração e generalização podem ser recursos utilizados por crianças no processo inicial de aquisição lexical.

Como será apresentado na seção sobre a natureza dessas habilidades (capítulo 3), há evidências da capacidade de abstração e generalização, como apontam trabalhos experimentais tanto em língua artificial com adultos (GÓMEZ, 2002; NEWPORT & ASLIN, 2004) quanto com bebês de sete e dezoito meses, respectivamente (MARCUS *ET AL.*, 1999; GÓMEZ, 2002). Decidimos estabelecer essa divisão na revisão da literatura existente, considerando que os trabalhos que serão apresentados no capítulo 3, além de discutirem a aquisição de DNAs, ampliam suas discussões para a natureza da habilidade de identificação de padrões.

Entretanto, como veremos, a maioria dos trabalhos utiliza línguas totalmente artificiais (GÓMEZ, 2002; NEWPORT & ASLIN, 2004; MARCUS *ET AL.*, 1999; GÓMEZ, 2002) e, quando utilizam língua natural, trabalham com crianças mais velhas (aos 18 meses, SANTELMANN AND JUSCZYK, 1998). O diferencial deste trabalho é que utilizamos uma língua em que mantivemos propriedades inerentes das línguas naturais, como os itens funcionais (presentes nas DNAs) e o contorno prosódico – diferentemente das línguas totalmente artificiais. É importante salientar que o reconhecimento de itens funcionais pode ser fundamental – ou, no mínimo, pode ter um papel facilitador – para a identificação das DNAs, assim como as fronteiras prosódicas. Dedicaremos as próximas subseções a esses elementos facilitadores na identificação das DNAs.

## 2.2- Elementos funcionais

Tradicionalmente, postula-se que o léxico de uma língua é formado por elementos pertencentes a categorias lexicais e funcionais e essa divisão parece ser básica em todas as línguas naturais. Categorias funcionais são concebidas como categorias do léxico que contêm informação acerca das propriedades gramaticalmente relevantes na língua, além de proverem a junção entre o léxico e a sintaxe. Os elementos dessas categorias (Agr, Infl, Compl e D) servem para prover uma estrutura sintática hierárquica, na qual elementos de categorias lexicais são incorporados numa derivação linguística, assim como para prover informação pertinente à referência a entidades e eventos.

Nos últimos anos, pesquisas, como a do Projeto Cartográfico (CINQUE & RIZZI, 2008), sugerem que o inventário dos elementos funcionais é muito maior do que tradicionalmente tem se considerado. O projeto cartográfico é a tentativa de mapear a estrutura sintática, dando visibilidade aos mais diversos núcleos funcionais. Nasce a partir da cisão do IP em diversas categorias, que leva, portanto à cisão do CP e do DP. A proposta inicial é que as sentenças e as expressões são formadas por uma estrutura lexical e uma estrutura funcional mais alta, ambas correspondendo a blocos elementares hierarquicamente organizados (CHOMSKY, 1986). Verifica-se que a estrutura funcional é constituída por mais de um núcleo. Essa abordagem assume hierarquias distintas das projeções funcionais dominando VP, NP, AP, PP, IP etc. Segundo essa proposta, a universalidade das línguas está relacionada ao tipo de núcleos e especificadores que as projeções funcionais envolvem, em seu número, e em sua ordem relativa. As línguas diferem no tipo de movimento que elas admitem ou no conteúdo do que elas visivelmente realizam em cada núcleo e especificador.

O projeto cartográfico de Cinque e Rizzi (2008) faz uma crítica à divisão clássica do léxico entre classes fechadas e classes abertas. Segundo esses autores, definir as classes fechadas como funcionais (em oposição a classes abertas como lexicais) é uma tarefa que se torna insustentável, principalmente quando analisamos os adjetivos e advérbios. Em muitas línguas, como por exemplo a língua Yoruba (CINQUE, 2006), os adjetivos constituem uma classe fechada e com número reduzido de elementos. Em relação aos advérbios (CINQUE, 1999), parece não haver uma tendência para advérbios de VP serem uma classe mais aberta ou mais numerosa do que a de advérbios do IP.

Ao considerar os domínios de pesquisa tradicionais, como a sintaxe formal, semântica lexical, aspecto, tempo e informação estrutural mais próximos um do outro, os itens de categoria funcional se relacionam com as interfaces, contribuindo para uma melhor compreensão das questões sobre significado e uso. Consequentemente, segundo Cinque e Rizzi (2008, p. 47), o inventário de núcleos funcionais se torna mais amplo e diversificado em relação ao assumido na divisão entre classes abertas e fechadas. É composto por elementos como: determinantes, complementizadores, preposições, morfemas de modo, tempo, aspecto, polaridade e voz, verbos auxiliares, cópulas e outros verbos que carecem de uma estrutura argumental clara, demonstrativos, quantificadores, numerais, palavras de grau, itens *Wh*, pronomes indefinidos, morfemas de caso, morfemas comparativos e superlativos, entre muitos outros. O mapeamento das estruturas sintáticas deixaria transparecer, de modo refinado, a complexidade contida nelas. Os autores defendem que, apesar de a estrutura ter sempre tido um papel central na gramática gerativa, foi a partir do programa minimalista que o mapa estrutural ganhou o foco dos estudos, provavelmente impulsionado pelo grande número de categorias funcionais identificadas e sugeridas na análise sintática nessa época e nos anos subsequentes. Trata-se, portanto, de uma lente microscópica posta sobre as categorias funcionais com o objetivo de reconhecê-las, discriminá-las e depreender sua posição hierárquica determinada pela GU.

De acordo com a teoria linguística gerativista chomskyana (Chomsky 1995 e obras posteriores), assumida nesta tese, os elementos do léxico são compostos por matrizes de traços fonológicos, semânticos e formais. Diferentemente dos itens lexicais, seus traços semânticos não atribuem papel temático e codificam traços gramaticais, isto é, traços que determinam relações e processos gramaticais, entretanto veiculam informação pertinente à referência ou à força ilocucionária da oração e caracterizam-se por prover posições estruturais relevantes para o comportamento sintático de categorias lexicais.

Segundo Corrêa (2001), do ponto de vista da psicolinguística, traços formais são aqueles que se realizam em termos do que é sistemático na língua e podem se referir a: distinções conceituais gramaticalmente relevantes e que ocorre de forma regular na língua (como distinções morfológicas de número e gênero presentes nos elementos funcionais, como afixos e determinantes); relação entre predicados e argumentos; possibilidade de alteração na ordem canônica em virtude do uso da língua no discurso. Embora alguns traços formais possam ser compartilhados entre as línguas, as línguas variam em relação às propriedades dos seus traços formais, mais especificamente em relação às propriedades dos traços formais das

categorias funcionais. Assim, a gramática a ser identificada pela criança está representada nos traços formais dos itens funcionais.

Dessa forma, para que a criança adquira sua língua materna, ela terá que identificar, na interface fônica, o que há de regular na língua, baseando-se na frequência, regularidade, padrões prosódicos e fonéticos disponíveis no fluxo da fala. A partir dessas pistas, a criança identificará os elementos funcionais da língua. A criança também detectará certos padrões regulares e sistemáticos presentes nos elementos funcionais, ou seja, perceberá as variações morfofonológicas nesses elementos, as quais, posteriormente, serão representadas como distinções morfofonológicas.

Os itens funcionais podem ser caracterizados perceptualmente, ou seja, podem ser interpretados, selecionados, distinguíveis e organizados a partir de sentidos sensoriais, como a audição. Sob uma análise descritiva, por conta de seu *status* linguístico de classe fechada, são altamente frequentes e, consequentemente, facilmente identificáveis na fala. Sob uma análise acústica, tendem a ter durações vocálicas mais curtas, amplitudes mais fracas e estrutura silábica simplificada em comparação a itens das categorias lexicais (Morgan, Shi, Allopenna, 1996). Analisando distribucionalmente, tendem a ocorrer nas extremidades das unidades prosódicas, permitindo, desde muito cedo, sua identificação no *continuum* linguístico e facilitando a segmentação do contínuo da fala em unidades menores. A seguir, apresentaremos evidências de sensibilidade a elementos funcionais.

### **2.2.1- Evidências de sensibilidade aos elementos funcionais**

Estudos revelam que as capacidades de distinção das diferenças acústicas entre itens funcionais e lexicais já estão presentes em bebês recém nascidos (Shi *et al*, 1999). Através da técnica experimental de sucção não nutritiva, em experimento realizado com bebês de 3 dias de vida, os pesquisadores observaram que os bebês reagiram significativamente à mudança do tipo de estímulo a que foram familiarizados. Bebês familiarizados com itens funcionais reagiram à apresentação de itens lexicais e vice-versa. Bebês que escutaram o mesmo tipo de

estímulo durante a familiarização e o teste (listas diferentes de itens lexicais nas duas fases ou listas diferentes de itens funcionais nas duas fases) não reagiram à mudança de lista. Esses resultados sugerem que bebês com poucos dias de vida são sensíveis a propriedades acústicas dos itens de sua língua que podem ser usadas, mais tarde, na identificação e distinção de itens funcionais e lexicais. Em estudo posterior, os pesquisadores identificaram que o reconhecimento de itens funcionais em uma língua natural parece ocorrer entre 8 e 13 meses (Shi, Werker e Cutler, 2003). Além disso, uma série de estudos tem explorado a percepção de bebês e crianças a itens funcionais.

Shi e colaboradores (2006) sustentam que a alta frequência dos itens funcionais propiciaria a segmentação do vocabulário no *continuum* de fala por bebês aos 11 meses de idade, contribuindo para a aquisição de elementos do léxico de uma língua. As crianças foram familiarizadas com pseudónomes precedidos tanto de itens funcionais reais (“the” – artigo definido em inglês) quanto por pseudoitens funcionais. A alta frequência de alguns itens funcionais (“the”) em relação a outros (“her”) na fala dirigida à criança também foi considerada. Assim, se a primeira sílaba – o item funcional da sequência de familiarização fosse reconhecida como familiar pelos bebês, o pseudonome com o qual é pareado poderia ser “aprendido” também em sua forma isolada, o que gerou, como resultado, um maior tempo de fixação do olhar quando aparecesse isoladamente na fase teste, sugerindo que a presença dos itens funcionais permite a segmentação do *continuum* sonoro. Hohle & Weissenborn (2003) chegaram a conclusões similares ao relatarem discriminação entre palavras funcionais e lexicais em crianças alemãs entre 7 e 9 meses expostas ao continuum sonoro.

Os determinantes, também pertencentes às categorias funcionais (categoria D), caracterizam-se por serem núcleo do Síntagma Determinante (DP), determinando propriedades referenciais de um nome. No campo da psicolinguística, trabalhos que focalizam, sobretudo, o tratamento perceptual dos dados de fala assumem tanto artigos quanto demonstrativos e possessivos como elementos da Categoria D. Será esse o caso de nosso estudo. Ainda que na maioria das línguas estudadas os determinantes compartilhem as mesmas propriedades acústicas de outros itens funcionais, esses itens se apresentam de modo diverso no português, se considerarmos os demonstrativos e possessivos, além dos artigos. Podem ser monossílabos (o/a, um, meu etc.), mas também dissílabos (uma, esse(a), nosso(a) etc.) e trissílabos (aquele(a)); ainda, apresentam vogais plenas, e não apenas reduzidas.

Em português brasileiro, Name (2003), investigando o processo de identificação do sistema de gênero por crianças adquirindo o português, sugere a sensibilidade à forma fônica

de artigos definidos e indefinidos e demonstrativos em crianças brasileiras aos 14 meses. Uchôa (2013) sugere que bebês aos 13 meses de idade parecem ser sensíveis à forma fônica dos determinantes e Teixeira (2013) sugere que a percepção e a identificação de subcategorias dos itens funcionais (determinantes e pronomes) no *continuum* sonoro facilitariam o reconhecimento das classes dos elementos lexicais (nomes e verbos) por crianças aos 13 meses. Tais resultados sugerem que apesar de crianças, em fase inicial de produção, produzirem poucos itens funcionais, já ocorre de forma efetiva a representação dos mesmos e a distinção dos itens lexicais.

Vimos no início deste capítulo que tínhamos como objetivo ressaltar dois aspectos principais das DNAs: i) serem constituídas por elementos funcionais ii) e se apresentam frequentemente em fronteiras de domínios prosódicos. Na próxima seção, dedicaremos à apresentação do segundo aspecto, caracterizando os constituintes prosódicos, em especial os sintagmas fonológicos, visto que DNAs se apresentam frequentemente nas fronteiras desses sintagmas.

### **2.3- Pareamento DNAs e constituintes prosódicos**

A fala é organizada em constituintes prosódicos e identificá-los no *continuum* sonoro é um passo para segmentá-la. Além disso, constituintes prosódicos são um ambiente linguístico reduzido para tratamento do *input* e, portanto, identificar elementos frequentes e relações de coocorrência, por exemplo, dentro desses espaços menores constituiria uma tarefa mais “fácil” do que entre eles. No que diz respeito às DNAs, sintagmas fonológicos, sintagmas entoacionais e grupo clítico são níveis de interface com a sintaxe, tornando-se, dessa forma, ambientes produtivos para DNAs.

A estrutura sonora da fala tem sido dividida em dois níveis: (i) no nível segmental, o qual diz respeito ao estudo dos fones e dos fonemas, são focalizadas as sequências de sons discretos e segmentáveis, cujas propriedades são conferidas a cada segmento. Cada som é composto por um feixe de propriedades acústicas/articulatórias, denominadas traços. Assim, os segmentos podem ser reconhecidos através dos valores absolutos dos traços que os

compõem e identificados por oposição paradigmática, sem considerar, necessariamente, os demais elementos do *continuum* sonoro, ou seja, cada segmento é analisado separadamente; (ii) no nível suprassegmental, o qual diz respeito às propriedades que se estendem por mais de um segmento, as propriedades da estrutura sonora não podem ser decompostas em unidades individualmente distintas, tendo valores relativos e não absolutos e seguindo a classificações próprias da Fonologia Prosódica. Na literatura, a prosódia, sendo associada a traços linguísticos e paralinguísticos, tem sido considerada um componente suprassegmental da fala (MORAES, 1998).

Há na literatura uma certa dificuldade em se apresentar uma única definição da prosódia (SHATTUCK-HUFNAGEL e TURK, 1996). A prosódia, de acordo com uma perspectiva linguística, se refere às variações de tom, intensidade e duração do continuum sonoro e é constituída pelos seguintes elementos: o acento, o ritmo, a velocidade da fala, a tessitura, a qualidade de voz e a entoação (CRYSTAL, 1994), essa última constituída por três parâmetros: a frequência fundamental, a intensidade e a duração. A seguir, apresentaremos a definição dos três parâmetros.

**A frequência fundamental** (ou  $f_0$ ), determina a frequência com que essas ondas ocorrem, ou seja, corresponde ao número de vezes que as pregas vocais abrem e fecham por segundo e é expressa em hertz (Hz). O termo frequência indica o número de ciclos (neste caso, vibrações) que as cordas vocais realizam em um segundo. A frequência fundamental, neste caso, é a primeira frequência produzida na glote (espaço entre as cordas vocais). Em relação à voz, ela está ligada às variações de altura (sons agudos e graves) e de intensidade (sons fortes ou fracos). O valor de  $f_0$  varia de acordo com o falante: homens – de 80 a 200 Hz; mulheres - de 200 a 300 Hz; crianças - de 400 a 500 Hz. É importante pontuar que os valores da frequência fundamental na Fala Dirigida à Criança (FDC) tendem a ser bem mais elevados do que na fala dirigida ao adulto. Além disso, podem ocorrer variações em um mesmo indivíduo, de acordo com seu estado emocional, a hora do dia e o valor pragmático dado ao discurso. Quanto maior a frequência fundamental de um som, mais agudo ele é percebido, portanto mais alto será o *pitch*. Quanto menor a frequência fundamental de um som, mais grave ele é percebido, portanto mais baixo será o *pitch*. Assim, o *pitch* é entendido como a propriedade auditiva que permite ao ouvinte dispor, em uma escala de grave e agudo, as variações de F0 de um som por ele percebidas.

Gussenhoven (2004) considera a frequência fundamental como o traço prosódico mais significativo na determinação do padrão entoacional de um enunciado.

A **duração** corresponde ao tempo de execução de um determinado intervalo. Esse intervalo pode corresponder a um único segmento fonético ou parte dele, a uma sequência de segmentos ou, ainda, à ausência de tais segmentos, expressa em pausas silenciosas, como na fronteira de sintagma fonológico da nossa atividade experimental (Atividade 1). Esse parâmetro tem importância fundamental na definição dos ritmos das diferentes línguas e é medido em unidades de tempo, tais como segundos ou milésimos de segundos.

Já a **intensidade** é determinada pela amplitude de vibração das cordas vocais, que varia de acordo com a pressão exercida sobre elas pelo ar na região subglótica. Em termos perceptuais, a energia presente nos sons é captada pelo ouvido humano, o que permite classificá-los em fortes ou fracos. É medida em decibéis (dB).

Os parâmetros acústicos de frequência fundamental, duração e intensidade, que foram analisados nos estímulos da atividade experimental 2, são de extrema importância para a delimitação dos constituintes prosódicos.

Os enunciados linguísticos, segundo Nespor e Vogel (1986) podem ser organizados em constituintes fonológicos dispostos hierarquicamente. Tais constituintes são definidos com base em regras de mapeamento, que consideram as informações oferecidas pelos demais componentes da gramática e em regras fonológicas, cuja aplicação/restrição no interior ou na fronteira de constituintes oferece evidências para a postulação dos domínios prosódicos trata particularmente dos domínios prosódicos. As autoras propõem uma hierarquia de constituintes fonológicos, a saber: a sílaba ( $\sigma$ ), o pé ( $\Sigma$ ), a palavra prosódica ( $\omega$ ), o grupo clítico (C), o sintagma fonológico ( $\phi$ ), o sintagma entoacional (IP) e o enunciado fonológico (U). Nespor e Vogel (1986) ressaltam que nem todas as línguas possuem necessariamente todos os níveis dessa hierarquia prosódica e ainda que hierarquia prosódica tem sua geometria definida a partir de quatro princípios, apresentados a seguir:

**Principle 1.** *A given nonterminal unit of the prosodic hierarchy,  $X^P$ , is composed of one or more units of the immediately lower category,  $X^{P-1}$ .*

**Principle 2.** *A unit of a given level of the hierarchy is exhaustively contained in the superordinate unit of which it is a part.*

**Principle 3.** *The hierarchical structures of prosodic phonology are n-ary branching.*

**Principle 4.** *The relative prominence relation defined for sister nodes is such that one node is assigned the value strong (s) and all the other nodes are assigned the value weak (w).*

(NESPOR & VOGEL, 1986, p.7)

E a formação de cada constituinte prosódico segue a uma regra que pode ser generalizada nos seguintes termos:

#### ***Prosodic Constituent Construction***

*Join into an n-ary branching  $X^p$  all  $X^{p-1}$  included in a string delimited by the definition of the domain of  $X^p$ .*

(NESPOR & VOGEL, 1986, p.7)

#### **2.3.1 - O sintagma fonológico ( $\phi$ )**

O sintagma fonológico, constituinte pesquisado por nós na atividade experimental 1, é considerado o constituinte prosódico mais importante para a interface sintaxe-prosódia, visto que as fronteiras deste tipo de constituinte coincidem não só com fronteiras de palavras, mas também com fronteiras sintáticas (NESPOR e VOGEL, 1986). Além disso, o sintagma fonológico é o constituinte que reúne um ou mais grupos clíticos e é formado por itens lexicais, caracteriza-se ainda pelo alongamento da tônica antes da fronteira e por um contorno melódico próprio (GOUT e CHRISTOPHE, 2006). Os princípios de formação do sintagma fonológico são assim definidos por Nespor e Vogel (1986):

#### ***Phonological Phrase ( $\phi$ ) Formation***

**a.  $\phi$  Domain:** *The domain of  $\phi$  consists of a C which contains a lexical head (X) and all Cs on its nonrecursive side up to the C that contains another head outside of the maximal projection of X.*

**b.  $\phi$  Construction:** *Join into an n-ary branching  $\phi$  all Cs included in a string delimited by the definition of the domain of  $\phi$ .*

**c.  $\phi$  Relative prominence:** *In languages whose syntactic trees are right branching, the rightmost node of  $\phi$  is labeled s; in languages whose syntactic trees are left branching, the leftmost node of  $\phi$  is labeled s. All sister nodes of s are labeled w.*

(NESPOR e VOGEL, 1986, p. 168)

Vários são os estudos que investigam a sensibilidade de bebês a fronteiras prosódicas. Christophe e colaboradores (1994) investigaram, através da técnica de sucção não nutritiva, se bebês franceses recém-nascidos eram capazes de distinguir entre sentenças que apresentavam uma fronteira de sintagma fonológico ( $\phi$ ) e sentenças que não apresentavam tal fronteira. Ao ouvirem os estímulos, os bebês, com três dias de vida, exibiram comportamentos que sugerem que, nessa idade, bebês franceses são capazes de perceber as propriedades acústicas que sinalizam uma fronteira de sintagma fonológico, tais como o alongamento pré-fronteira e o contorno melódico, percebido como variações de *pitch*. Além disso, Myers e colaboradores (1996) sugerem que, a partir dos onze meses, bebês são capazes de perceber fronteiras de palavras.

Hirsh-Pasek et al. (1987) constataram que bebês, com quatro meses de idade, reagem a rupturas nos domínios de sintagma entoacional. Segundo Jusczyk e colaboradores (1992), a partir dos nove meses de idade, bebês adquirindo o inglês são sensíveis às fronteiras de sintagmas sintáticos que correspondem a fronteiras de constituintes prosódicos. Os resultados de Silva (2014) sugerem que bebês brasileiros com idade média de treze meses são sensíveis às propriedades prosódicas que demarcam fronteiras de sintagma entoacional (I) na Fala Dirigida à Criança (FDC) e as utilizam como pistas para a segmentação do *continuum* da fala.

Aos nove meses, bebês exibem sensibilidade a rupturas nos domínios de sintagmas fonológicos (KEMLER-NELSON et al, 1989). Gout et al. (2004) investigaram se o mesmo tipo de fronteira restringiria o acesso lexical durante o processo de aquisição da linguagem, sugerindo que esta pista prosódica é usada por bebês de 10 a 13 meses adquirindo o inglês.

## 2.4- Conclusão

Buscou-se, neste capítulo, caracterizar as DNAs, foco do nosso estudo, ressaltando duas propriedades das mesmas: serem constituídas por elementos funcionais e se apresentam frequentemente em fronteiras de domínios prosódicos, como o sintagma fonológico.

Foram apresentadas as características, propriedades e importância das DNAs na aquisição de uma língua natural, bem como uma revisão bibliográfica sobre a sensibilidade de bebês à padrões de DNAs. Vimos que a grande maioria dos estudos apresentados investigam as relações de DNAs em línguas que não apresentam as mesmas (ou ao menos algumas) características de línguas naturais (línguas artificiais) e que, quando essas relações são investigadas em línguas naturais, não podemos assumir que houve a abstração e generalização de um novo padrão, e sim um reconhecimento de um padrão já existente na língua materna da criança.

Como vimos, as dependências não adjacentes são constituídas, em sua grande maioria, de itens funcionais e, por essa razão, buscou-se explorar um pouco mais a literatura em aquisição da linguagem pertinente aos elementos funcionais, de acordo com a concepção da teoria linguística gerativista chomskyana. Uma série de estudos foram apresentados evidenciando que crianças são capazes de identificar e representar os elementos funcionais no *continuum sonoro*.

Buscou-se ainda apresentar as características dos constituintes prosódicos, visto que constituem um ambiente linguístico reduzido para tratamento do *input* e sintagmas fonológicos, sintagmas entoacionais e grupos clíticos são níveis de interface com a sintaxe, tornando-se, dessa forma, ambientes produtivos para DNAs. Foram apresentados alguns estudos que sugerem que bebês são sensíveis à fronteiras prosódicas ainda nos primeiros dias de vida.

Em nosso estudo, ao assumirmos bebês nascem com uma predisposição biológica para a aquisição de uma língua natural, assumimos também que o *input* linguístico apresenta propriedades que servirão como pistas – como a relação de DNAs entre itens funcionais, e ambientes prosódicos específicos – como fronteiras de sintagma fonológico, que facilitarão a aquisição lexical.

A partir das propriedades das DNAs apresentadas e de estudos realizados anteriormente (TEIXEIRA, 2012), apresentamos como objetivos específicos investigar se i) DNAs podem facilitar a categorização inicial e se ii) um ambiente prosódico específico – extremidades do sintagma fonológico - pode facilitar o mapeamento, abstração e generalização de DNAs, ou seja, investigar se alguns aspectos da fonologia podem ser explorados para facilitar a identificação de relações morfossintáticas.

Assim, em nossa atividade experimental 1, buscamos investigar se a presença de DNAs alinhadas à fronteira de sintagma fonológico seria um facilitador para a abstração e generalização do padrão de DNA estabelecido entre determinante e terminação de nome. E, em nossa atividade experimental 2, buscamos investigar se a identificação, abstração e generalização de DNAs entre Det e terminação de N, poderia facilitar a categorização de itens do léxico.

Como vimos na seção 2.2.1, há vários estudos que se dedicam à investigação do reconhecimento e diferenciação entre itens lexicais e funcionais desde tenra idade. Em nosso estudo, optamos por ir além dessa investigação, focalizando as DNAs, visto que é um fenômeno altamente frequente em língua natural e, como vimos, um “ambiente” tipicamente composto por itens funcionais. Ao elegermos as DNAs como foco de estudo, não estamos investigando somente o tratamento estatístico de palavras mais frequentes ou a distinção entre elementos funcionais e lexicais, mas buscamos investigar o reconhecimento, abstração e generalização da relação entre esses elementos. Para tal, assumimos que a identificação dessa relação, por parte do aprendiz, pode ser considerada mais que um tratamento estatístico, pois não trata-se mais apenas da identificação de frequência dos itens. Acreditamos que essa capacidade já faz parte ou depende de uma habilidade computacional, um tratamento simbólico das informações, permitindo o mapeamento e a abstração das relações presentes, como veremos no capítulo 3.

### **3- HABILIDADES DE SEGMENTAÇÃO E IDENTIFICAÇÃO DE PADRÕES**

Neste capítulo, serão apresentadas as habilidades que estamos assumindo como disponíveis para os bebês na identificação e segmentação de padrões do *continuum* linguístico e a discussão sobre a natureza dessas habilidades, ou seja, se elas seriam habilidades de domínio geral ou de domínio específico. Buscaremos avaliar em que medida essas teorias podem contribuir e dialogar com nosso trabalho. Após a caracterização dessas habilidades, será apresentada a proposta de integração (GERVAIN & MEHLER, 2010), assumida nesta tese, entre mecanismos estatísticos e habilidades computacionais.

#### **3.1- O Bootstrapping prosódico**

De acordo com a proposta do *bootstrapping* prosódico, uma análise fonológica do sinal da fala pode permitir às crianças adquirirem o léxico e a sintaxe de sua língua. Essa hipótese assume que a criança não precisa ter conhecimento linguístico prévio e, por isso, se apresenta como uma boa proposta que investiga as etapas iniciais da aquisição.

Christophe *et al.* (1997) apresentam quatro tipos de informações utilizadas pelas crianças para identificarem, no *continuum* da fala, as formas de palavras. É importante ressaltar que as informações apresentadas em i, iii e iv estão presentes nos estímulos de nossas atividades experimentais. São elas:

- i) regularidades distribucionais: posição fixa e recorrente de determinados itens em sintagmas e frases;
- ii) regularidades fonotáticas: probabilidade de ocorrência contígua de determinados fonemas;
- iii) formas típicas de palavras: características inerentes e específicas às categorias funcionais e às categorias lexicais;
- iv) pistas de fronteiras prosódicas.

Essa hipótese sugere que, a partir da análise do sinal acústico da fala, a criança reconheceria padrões fonotáticos, distribucionais e prosódicos de sequências fonológicas, o que viabilizaria o acesso à estrutura sintática e aos itens lexicais de sua língua. Regularidade distribucional diz respeito à intuição por parte da criança de que sequências sonoras que

ocorrem com frequência em uma variedade de contextos são melhores candidatas ao léxico do que as que ocorrem raramente em poucos contextos. Informações fonotáticas referem-se às limitações na ocorrência de sons entre palavras e sentenças, considerando que alguns sons ou combinações de sons ocorrem somente em fronteira de palavra, e outros ocorrem apenas no interior da palavra.

A partir das propriedades prosódicas e distribucionais dos enunciados linguísticos e com base em evidência experimental relativa à sensibilidade de bebês a essas propriedades, Christophe e colaboradores (1997, 2008) propõem um modelo das etapas iniciais da aquisição lexical e sintática, cujo esquema se encontra abaixo:

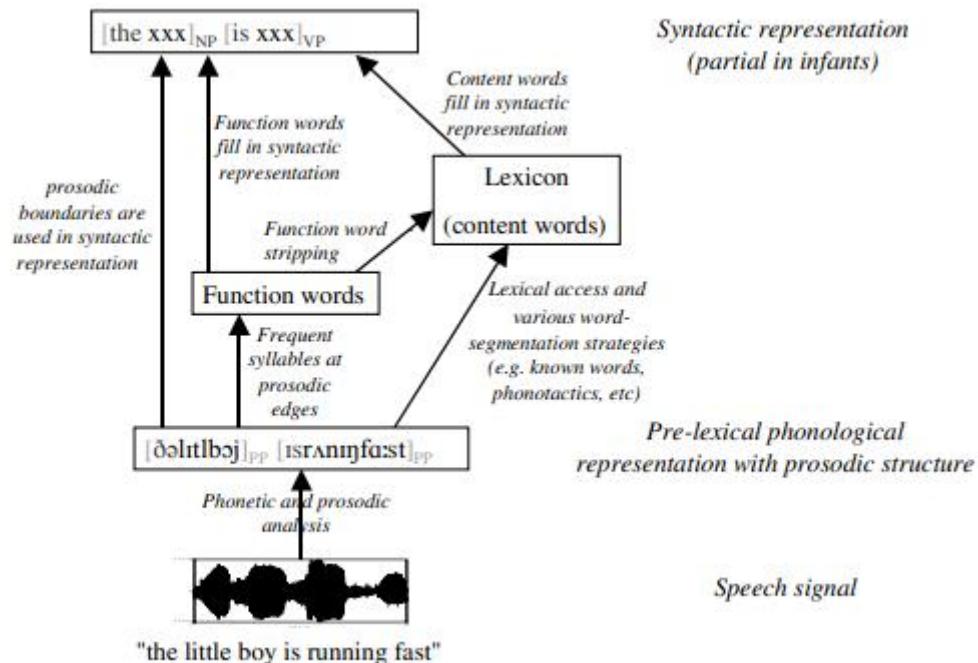


Figura 1 – Modelo dos estágios iniciais do processamento de fala (Christophe *et al.*, 2008)

De acordo com o modelo esquematizado acima, a análise acústica permite - via prosódia - a segmentação em unidades menores e essas em unidades ainda menores como palavras ou ainda elementos menores que as palavras, gerando uma representação pré-lexical (sem conteúdo semântico). Nessa segmentação, os itens funcionais, devido à sua alta frequência no estímulo, e informações prosódicas são destacados, facilitando a identificação dos itens lexicais. Itens funcionais e itens lexicais são mapeados por computação simbólica, o que poderíamos chamar de um pré-processamento sintático, o qual contribui para a aquisição lexical.

A hipótese do *bootstrapping* prosódico busca, portanto, dar conta de alguns questionamentos básicos que envolvem o processo de aquisição. Primeiramente, busca investigar quais informações são fornecidas pelo sinal de fala que não dependem do conhecimento sintático/lexical para serem identificadas. Ainda busca investigar (i) se os bebês têm habilidades para reconhecer e usar tais pistas; (ii) quais habilidades de tratamento das propriedades acústicas do sinal de fala os bebês dispõem e (iii) a partir de quando essas habilidades estão disponíveis. A hipótese do *bootstrapping* prosódico torna-se, portanto, alicerce para a pesquisa desenvolvida nesta tese, ao assumirmos que mecanismos de abstração e generalização de padrões do sinal acústico são recursos disponibilizados precocemente e explorados pelos bebês de modo a levar à categorização de elementos lexicais, mesmo expostos a um pequeno número de enunciados. Assim, diante de um *continuum* sonoro linguístico há a segmentação do fluxo da fala em unidades cada vez menores e a distinção entre elementos lexicais e funcionais a partir da identificação e generalização de padrões – distribucionais, fonotáticos e prosódicos –, viabilizando a abstração e generalização de padrões frequentes de DNAs e, posteriormente, a aquisição da estrutura sintática e, consequentemente, do léxico de sua língua materna.

### **3.2- Habilidades de identificação de padrões**

Nas últimas décadas, eclodiram os estudos que tratam das habilidades tanto de bebês quanto de adultos na identificação de padrões linguísticos. De forma geral, os trabalhos sustentam o uso somente de habilidades estatísticas ou o uso de habilidades estatísticas e computacionais/simbólicas na identificação dos padrões linguísticos.

A fim de conduzir a leitura da revisão bibliográfica que se segue, faz-se necessário salientar que primeiro apresentaremos, em ordem cronológica, os estudos que consideram somente o uso de habilidades estatísticas (Saffran et al., 1996; Gómez, 2002; Gómez & Lakusta, 2004; Gómez & Maye, 2005; Van den Bos et al., 2012; Langus et al., 2012) e, em um segundo momento, também em ordem cronológica, os estudos que defendem a integração entre mecanismos gerais e mecanismos computacionais/simbólicos (Marcus et al., 1999; Peña et al., 2002).

Investigando as habilidades de identificação de padrões e assumindo habilidades estatísticas no tratamento das informações, vários são os estudos que utilizam línguas artificiais.

Saffran e colaboradores (1996), pioneiros nestes estudos, defendem que crianças ainda muito novas são capazes de utilizar informação estatística para segmentar o *continuum sonoro*. Neste estudo, os autores investigaram o direcionamento do olhar de crianças aos 8 meses. Esse direcionamento estaria relacionado à informação estatística inerente nas sequências de sentenças produzidas em língua artificial, como, por exemplo, probabilidades transicionais<sup>3</sup> nas quais é estimada a probabilidade de um item seguir imediatamente outro, sugerindo que habilidades que levam em conta informações similares (informação estatística) podem ser consideradas mecanismos facilitadores da aquisição de uma língua. Após esse estudo precursor, uma gama de outros estudos surgiu investigando o papel, o escopo e as limitações do aprendizado estatístico na aquisição da linguagem.

Gómez (2002), com o intuito de investigar o processo de reconhecimento e generalização de dependências não-adjacentes (DNA) e, considerando que uma possível base para essa aprendizagem seriam as dependências adjacentes (DA) (condicionantes de ordem inferior numa hierarquia em que dependências adjacentes seriam aprendidas anteriormente e serviriam de base para o aprendizado de DNA), investigou a possibilidade de aprendizagem de DNA por adultos e bebês aos 18 meses. Para a realização das tarefas, duas línguas artificiais com 3 elementos distintos foram elaboradas de modo que os estímulos nas duas línguas começavam e terminavam com os mesmos elementos e continham os mesmos elementos de transição (**x**) que representam o número da variabilidade do elemento intermediário, que poderia ser variável em 2, 6, 12 ou 24 elementos. Por exemplo, língua 1: “pel-**x**-rud”, “vot-**x**-jic”, “dak-**x**-tood”; língua 2: “pel-**x**-jic”, “vot-**x**-tood”, “dak-**x**-rud”. Dessa forma, os aprendizes poderiam distinguir as línguas somente pela aquisição das DNAs. Além disso, a maior variabilidade de **x** propiciaria uma maior focalização nas DNA e, consequentemente, um aprendizado mais eficaz das mesmas. Após 18 minutos de exposição e agora já na fase teste, os adultos eram informados que os estímulos ouvidos durante o treinamento pertenciam a um conjunto de regras de ordem de palavras e que ouviriam, nesta fase, mais 12 estímulos (6 com a mesma estrutura da fase de familiarização + 6 novos - pertencentes ao outro grupo da língua artificial). Eram também instruídos a selecionar

---

<sup>3</sup> Um exemplo de probabilidade transicional em língua natural é a sílaba “pre” em inglês. “Pre” precede um conjunto reduzido de sílabas, incluindo “ty”, “tend” e “cedes”. No continuum linguístico, a probabilidade de “pre” ser seguido por “ty” é relativamente alta (cerca de 80% na fala dirigida à criança). Entretanto, devido à sílaba “ty” ocorrer em finais de palavras, pode ser seguida por qualquer sílaba que inicie uma palavra em inglês. Dessa forma, a probabilidade de “ty” ser seguida por “ba”, como em “pretty baby” é extremamente baixa (cerca de 0.03% na fala dirigida à criança). Essa diferença na probabilidade sequencial é o que determina que “pretty” é uma palavra da língua inglesa e “tyba” não é (Saffran, 2003).

determinada tecla em um computador se os estímulos ouvidos fossem congruentes com os ouvidos anteriormente. Os resultados obtidos com adultos sugerem que o aumento de variabilidade do elemento intermediário propicia um aprendizado mais eficaz das DNAs. Assim, participantes expostos ao conjunto de 24 variações para “x” apresentaram resultados mais satisfatórios que os dos outros grupos, evidenciando melhor aprendizado de DNAs. A fim de se verificar a mesma habilidade em crianças e, ainda, se o aumento da variabilidade do elemento intermediário também facilitaria a identificação de padrões de DNAs, bebês, aos 18 meses, foram expostos aos mesmos estímulos, com período de familiarização de 3 minutos e testados com a técnica do Olhar Preferencial. Os resultados sugerem que o desempenho dos bebês, assim como o de adultos, aumentou radicalmente com o aumento da variabilidade do elemento intermediário, ou seja, bebês expostos ao conjunto de 24 variações para “x” também apresentaram resultados mais salientes que os dos outros grupos. Ainda que seja um pouco forçado comparar os resultados das duas atividades experimentais, principalmente devido à diferença do tempo de exposição à língua artificial (18 min (adultos) e 3 min (bebês)) e pelo fato de que os adultos foram instruídos para a realização da tarefa, os resultados foram ao encontro da hipótese inicial de que a alta variabilidade do contexto poderia levá-los a concentrarem-se em estruturas menos variáveis, ou seja, DNAs. De modo mais amplo, os resultados sugerem que aprendizes são sensíveis aos elementos constituintes das DNAs.

Gómez e Lakusta (2004), com o objetivo de investigar como crianças começam adquirir categorias baseadas na forma e nas relações entre elas, ou seja, se as crianças seriam capazes de identificar, abstrair e generalizar o padrão da língua às novas formas, propuseram duas atividades experimentais expondo crianças de 12 meses à estrutura sonora da forma aX e bY (correspondendo “a” e “b” a elementos de categorias funcionais e “X” e “Y” a lexicais, como em uma língua natural). Segundo as autoras, assimetrias categóricas desse tipo podem desempenhar um importante papel na aprendizagem, destacando pontos de ancoragem para a análise distribucional. De forma sucinta, o objetivo do estudo era identificar se crianças de 12 meses poderiam estabelecer as relações entre funções de palavras “a” e “b” e categorias X e Y baseadas em seus traços distintivos, a fim de generalizá-las a pares de palavras “a” e “b” com novos elementos “X” e “Y”. Os resultados sugerem que crianças, aos 12 meses, são capazes de discriminar pares de itens funcionais e lexicais congruentes de incongruentes a partir da distinção das pistas presentes.

Gómez & Maye (2005) também investigaram o aprendizado de DNAs em língua artificial com bebês de 12 e 15 meses, replicando a atividade de Gomez (2002) em que a alta variabilidade do elemento interveniente contribui para a percepção da DNA. Os resultados

indicam que aos 12 meses e sob as condições experimentais apresentadas, os bebês não foram capazes de rastrear as DNAs, diferentemente dos bebês aos 15 meses.

Investigando as habilidades de identificação de padrões de DNA em adultos, Van den Bos et al (2012) e Langus et al. (2012) investigaram se falantes adultos poderiam rastrear DNA na presença de outras pistas. O primeiro estudo conciliou DNAs com pistas visuais e fonológicas, enquanto o segundo investigou se adultos podem simultaneamente rastrear variações de *pitch* e duração de sílaba ao segmentar o continuum linguísticos e agrupá-las em sentenças.

O conjunto de estudos apresentados evidencia que algumas pesquisas, além de utilizarem língua totalmente artificial, sustentam que habilidades estatísticas e probabilísticas, de domínio geral, subjazem ao tratamento de DNAs por bebês e adultos (SAFRAN et al. 96; GÓMEZ, 2002; GÓMEZ & MAYE, 2005; VAN DEN BOS et al 2012 e LANGUS et al. 2012).

Outros estudos (Marcus e colaboradores, 1999; Peña e colaboradores 2002) defendem que habilidades estatísticas, sozinhas, não dão conta do tratamento de DNAs em língua natural e, portanto, bebês recorrem tanto a mecanismos gerais quanto a mecanismos computacionais/simbólicos. De acordo com esses trabalhos, mecanismos estatísticos podem ser suficientes para o reconhecimento/rastreamento de elementos que se apresentam em configurações de DA e DNA, mas para o uso produtivo dessas configurações – abstraindo um padrão e generalizando-o em novos enunciados, a criança ou adulto precisaria de habilidades computacionais, simbólicas.

Marcus e colaboradores (1999) propõem ao menos 2 mecanismos de aprendizagem: mecanismo estatístico (Saffran *et al*, 1996) e regras algébricas. Assumindo a hipótese de que crianças são capazes de abstrair a estrutura adjacente de uma língua e generalizá-la a novas instâncias, os autores esperam que crianças, utilizando o paradigma de escuta preferencial, atentem por mais tempo, na fase teste, a itens inconsistentes em relação aos itens consistentes.

Nas três atividades experimentais, conduzidas em bebês de 7 meses de idade eram habituados, durante 2 minutos, a sentenças de três palavras construídas a partir de uma língua artificial (no formato “ABA” ou “ABB”) e, posteriormente, testadas com novas sentenças de três palavras, as quais não haviam aparecido na fase de habituação. Devido ao fato de as sentenças testes serem novas, em relação às da fase de habituação, as crianças não poderiam distinguir as sentenças baseadas nas probabilidades transicionais, assim como por serem constituídas de mesma duração e contorno prosódico elaborado em computador, não era possível distingui-las baseando-se apenas nas propriedades estatísticas, como número de

sílabas. Os resultados da primeira atividade experimental evidenciaram que as crianças mostraram preferência por sentenças inconsistentes em relação às consistentes, como era esperado a partir do pressuposto de preferência pela novidade.

Entretanto, devido ao fato de a sequência dos traços fonéticos das sentenças do teste ser próxima à da habituação, uma segunda atividade experimental foi proposta. Diferentemente da primeira, os itens da fase de habituação não proveriam informações sobre a relação entre consoantes vozeadas e desvozeadas, distinguindo os traços fonéticos das fases de habituação e teste. Mais uma vez, os resultados foram de acordo com a previsão, ficando as crianças atentas por mais tempo aos itens inconsistentes em relação aos consistentes. A partir da possibilidade de que crianças poderiam ter se habituado com a propriedade de reduplicação que distingue as duas “gramáticas” “ABA/ABB” (repetições em ABB contêm elementos imediatamente reduplicados (“ti ti”), o que não ocorre nas sequências de “ABA”), uma terceira atividade experimental foi proposta - comparação entre sentenças construídas pela formação “ABB” e sentenças construídas pela formação “AAB” com os mesmos itens do experimento 2 – e os resultados foram também compatíveis com a previsão.

A partir dos resultados, os autores propõem um mecanismo pelo qual crianças extraem um tipo de regra abstrata que representem relações entre variáveis, como “o primeiro item de X é o mesmo do terceiro item de Y”. Assim, os resultados parecem mostrar que crianças possuem a capacidade de extrair rapidamente essas regras de um pequeno número de ensaios e de generalizar tais regras para novas instâncias. Segundo os autores, se essa posição estiver correta, o bebê, nos primeiros meses de vida, possui no mínimo duas ferramentas distintas de aprendizagem: uma que trata de relações estatísticas, como probabilidades transicionais, e outra, que manipula variáveis, permitindo o aprendizado de regras por crianças.

Seguindo a mesma proposta, Peña e colaboradores (2002) propõem que o aprendizado de uma língua requer um mecanismo estatístico para identificar as palavras no fluxo da fala e computações algébricas para “descobrir” o mais alto nível estrutural (gramatical). Na atividade desenvolvida com adultos foi criado um continuum de pseudolíngua que possuía elementos trissílabos caracterizados por suas probabilidades transicionais não adjacentes. Os elementos da “língua AXC” poderiam aparecer com três diferentes Xs, criando uma família de palavras, como [puliki], [puraki] e [pufoki]. Assim, após a fase de familiarização, foram apresentados aos participantes palavras (AXC) e partes de palavras (CAX e XCA). O conjunto de resultados sugere que adultos são capazes de segmentar o fluxo da fala, quando micro pausas – consideradas próximas a uma pista prosódica- foram inseridas, usando apenas

o mecanismo estatístico, mas não são capazes de extrair regularidades estruturais presentes no continuum mesmo quando o tempo de familiarização é maior.

A hipótese que norteia este trabalho está em consonância com a proposta de Marcus et al. (1999) e Peña et al. (2002) de que habilidades estatísticas, sozinhas, não dão conta do tratamento de DNAs em língua natural e, portanto, bebês recorrem tanto a mecanismos gerais quanto a mecanismos computacionais/simbólicos, especificamente linguísticos, na aquisição de uma língua.

Dessa forma, a fim de propor soluções para o problema da identificação de padrões presentes no *continuum* linguístico, várias hipóteses surgem com o intuito de elucidar a questão de como os falantes adultos e, nesta tese, os bebês, são capazes de segmentar o *continuum*, identificar padrões e extraí-los. Tais padrões servirão de suporte inicial para aquisição de sua língua materna. Assumimos que existem pistas variadas, como pistas distribucionais, pistas de fronteiras prosódicas, regularidades fonotáticas, ocorrência de marcadores morfossintáticos em ambientes frequentes, dentre outras pistas. Dessa forma, nesta tese, acreditamos que o conjunto de todas as pistas, quando disponíveis no estímulo, juntamente com a habilidade estatística e uma habilidade de tratamento simbólica seria responsável pelo processo de segmentação do contínuo sonoro linguístico.

### **3.3- Habilidades de domínio geral x Habilidades de domínio específico**

Antes de apresentarmos a proposta de integração entre tratamento estatístico e habilidades computacionais no tratamento do estímulo linguístico assumida neste trabalho, apresentaremos a dualidade entre *habilidades de domínio geral* e *habilidades de domínio específico* e discutiremos como essas teorias contribuem e dialogam com nosso trabalho.

Skinner<sup>4</sup> expôs, com a publicação em 1957 de *Verbal Behavior*, sua teoria behaviorista de aquisição da linguagem, assumindo mecanismos equipotenciais para a aquisição da

---

<sup>4</sup> Enquanto o psicólogo comportamentalista Skinner propõe em seu livro “Verbal Behavior” (1957) que a linguagem, como qualquer comportamento animal, é um operante que viria a se desenvolver na criança através do reforço que ela recebe do meio externo, Chomsky, alguns anos após, defende a existência de uma “Faculdade da linguagem” que inclui restrições inatas específicas que uma possível língua a ser aprendida precisaria tomar. Dessa forma, para os behavioristas, aprender uma língua seria igual a aprender qualquer tarefa cognitiva, humana ou não. Assim como aprendemos uma segunda língua, aprender a língua materna, de acordo com essa visão, seria o resultado de uma instrução direta sobre as regras dessa língua. Diferentemente, segundo a proposta inatista de Chomsky, adquirir uma língua envolve um processo subconsciente das regras gramaticais que estão dispostas nos estímulos aos quais a criança é exposta.

linguagem através da detecção de contingências entre entidades observáveis. Dessa forma, o aprendizado de uma língua era explicado com base em experiências e reforços de cada organismo, através dos mesmos mecanismos observados na aprendizagem de outros domínios e em outras espécies. A teoria do desenvolvimento de Jean Piaget (1983), por exemplo, assim como abordagens que defendem a aprendizagem estatística são tipicamente consideradas teorias de mecanismos de domínio geral.

A transição do foco de teorias da aprendizagem gerais para teorias de aprendizagem que defendem a especificidade de domínio começou em meados do século XX (MAYER, 2002). Chomsky, um dos principais nomes desse movimento, altera radicalmente o campo de conceptualização no qual a aquisição da linguagem se enquadra, expondo a necessidade de um mecanismo de aprendizagem mais especializado ao assumir que uma gramática internalizada permite aos aprendizes irem além dos dados do *input*, permitindo generalizações. A crítica feita por Chomsky (1959) à proposta de Skinner defende, em um de seus pontos principais, que representações internas são necessárias para explicar o funcionamento da linguagem.

The fact that all normal children acquire essentially comparable grammars of great complexity with remarkable rapidity suggests that human beings are somehow specially designed to do this, with data-handling or ‘hypothesis-formulating’ ability of unknown character and complexity (Chomsky, 1959).

O resultado dessa transição foi o surgimento de discussões que se concentram em como os aprendizes se desenvolvem e aprendem sobre áreas específicas (MAYER, 2002). Enquanto teorias que propõem a generalidade de domínio defendem que o pensamento de uma criança em uma dada idade pode ser caracterizado em termos de um único nível cognitivo, teorias que defendem a especificidade de domínio argumentam que sistemas cognitivos “descobrem”/acessam propriedades distribucionais do *input*.

“Habilidades cognitivas são de domínios específicos na medida em que o modo de raciocínio, a estrutura de conhecimento e os mecanismos para adquirir conhecimento diferem em importantes modos entre as distintas áreas de conteúdo” (GELMAN, 1999, p. 238, tradução nossa). Uma série de evidências para a especificidade de domínio foram apresentadas na literatura, entre as quais destacam-se: i) a variabilidade de um nível cognitivo entre domínios em um determinado indivíduo e em um determinado ponto no tempo,

sugerindo a não possibilidade de uma habilidade de domínio geral, visto que, caso assim fosse, ocorreria no mesmo momento em todos os indivíduos (GELMAN e BAILLARGEON 1983 *apud* GELMAN, 1999); ii) capacidades cognitivas inatas em recém-nascidos, as quais possuem ao menos duas características principais: serem de domínio específico e específicos por tarefa (SPELKE, 1994). Segundo Spelke (1994), crianças parecem apresentar conhecimento sistemático em quatro domínios: física, psicologia, números e geometria. Entretanto, a principal evidência de que a cognição humana é construída sobre sistemas de conhecimento de domínio específico é proveniente de estudos sobre o raciocínio de crianças no domínio da física. Nesse domínio, crianças bem novas parecem fazer inferências sobre os movimentos ocultos de objetos inanimados de acordo os princípios de coesão, continuidade e contato<sup>5</sup>; iii) evidências etológicas de aprendizagem animal, sugerindo, por exemplo, um paralelo entre a aprendizagem do canto em aves e a aquisição da fala em humanos, nos quais a escolha do que aprender é guiada pelas preferências inatas generalizadas, resultando em aprendizados específicos de tradições vocais (Marler, 1991; Pinker 1994).

De acordo com Gelman (1999), a concepção de mente em domínios específicos de aprendizagem se apresenta em várias teorias, dentre as quais se destaca a *teoria da modularidade* de Fodor, a qual propõe que a mente é formada por vários módulos de processamento de informação e esses módulos operam de forma relativamente independente uns dos outros, processando somente um tipo específico de informação, seja ela corporal, visual, auditiva, linguística ou de qualquer outra esfera.

Fodor (1983) propõe uma caracterização da estrutura da mente considerando os diferentes tipos de representações e processos cognitivos. Baseado na filosofia do neocartesianismo, no pensamento do frenologista Franz Gall (1758- 1828) e no pensamento de Chomsky (1980), postula uma teoria de arquitetura mental, segundo a qual os processos mentais constituiriam módulos independentes ou encapsulados, cada qual funcionando com regras e processos próprios, organizados em módulos verticais. Ele distinguiu entre processos cognitivos centrais e sistemas perceptuais, defendendo a modularidade destes últimos. Fodor defende que os processos mentais são melhor caracterizados como módulos independentes ou “encapsulados”, funcionando cada um com regras próprias e processos próprios em “módulos

---

<sup>5</sup> Segundo Spelke (1994), no princípio de coesão, “Um objeto em movimento mantém a sua conexão e limites”; no princípio de continuidade, “Um objeto em movimento traça exatamente um caminho conectado ao longo do espaço e do tempo” e no princípio de contato “Os objetos se movem em conjunto, se e somente se eles se tocarem”.

verticais”, como a linguagem e a visão, cada um com seu modo de operação característico. Esses módulos estariam ligados a um “processador central” que teria acesso às informações de todos os módulos, além de informações armazenadas na memória de longo prazo. O autor defende, dessa forma, que o processador seja “flexível” para tomar decisões e resolver problemas por meio dos vários *inputs* que recebe, permitindo que se construam crenças de como o mundo é.

Muitos são os pesquisadores que sugerem que os modos pelos quais a linguagem é adquirida e representada são distintos dos modos pelos quais outras habilidades cognitivas são aprendidas e representadas (Chomsky, 1988; Pinker, 1994; GLEITMAN, 1994). Outros domínios incluem (mas não se limitam) ao processamento de números, percepção de faces e raciocínio espacial. No que tange à linguagem, e, mais especificamente, à sintaxe – principal foco de estudo de Noam Chomsky - podemos considerar como evidências para a sintaxe ser considerada um módulo, sua inatidão, seu caráter biológico (exclusiva em humanos), sua localização neurológica e degradação, sua aquisição rápida diante de dados escassos e a presença de períodos críticos e tempos de maturação (ver Pinker, 1994). De acordo com Gelman (1999), as teorias sobre modularidade, de forma geral, variam em ao menos dois aspectos: (i) se a modularidade é restrita aos processos perceptivos ou afeta os processos que envolvem raciocínio, e (ii) se modularidade é inata ou construída. Assumir a modularidade não implica assumir, necessariamente, módulos inatos envolvidos; entretanto, a maioria dos autores que defendem a modularidade também defendem o inatismo dos módulos. É concenso, no entanto, que todas as visões sobre modularidade assumem a especificidade de domínio.

Como vimos, teorias de aprendizagem que defendem a especificidade de domínio contrastam com teorias que assumem que os seres humanos são dotados de um conjunto geral de habilidades de raciocínio (como a memória, a atenção e a inferência) que se aplicam a qualquer tarefa cognitiva, independentemente do conteúdo específico. Generalidade de domínio implica um conjunto de mecanismos de aprendizagem simples e generalizados que apresenta seu funcionamento restrinido a certos tipos de *input*, mas não a outros, como uma função da percepção e cognição humana, enquanto que especificidade de domínio geralmente implica um sistema de conhecimento modular e inato, ainda que modularidade possa emergir como uma função da experiência dentro de um domínio particular.

Saffran & Thiessen (2007)<sup>6</sup> avaliam em que medida as capacidades de aprendizagem de domínio-geral podem contribuir para a aquisição das línguas naturais. Ainda que a utilização e integração de múltiplas pistas pelos bebês possam ocorrer a partir de processos de domínio geral – como defendem Saffran e Thiessen (2007) – a segmentação, por exemplo, parece ser apenas uma das etapas do processo de aquisição lexical, visto que o bebê terá ainda como tarefa a extração de propriedades distribucionais e morfofonológicas que o “guiarão” à aquisição lexical. Ou seja, o estímulo linguístico parece, de fato, exigir um tratamento diferenciado em relação aos de outros domínios.

Nos últimos anos, vários autores têm defendido o uso de habilidades estatísticas por bebês na aquisição de uma língua (SAFRAN et al. 96; GÓMEZ, 2002; GÓMEZ & MAYE, 2005; VAN DEN BOS et al. 2012 e LANGUS et al. 2012). Como vimos, essas habilidades também seriam utilizadas em outros domínios, caracterizando-as como habilidades de domínio geral. Entretanto, no que diz respeito ao domínio linguístico, sozinhas, essas habilidades não seriam suficientes para explicar a aquisição do que estaria subjacente à estrutura da língua e o modo como as crianças abstraem e generalizam tais relações estruturais. Surge, dessa forma, uma proposta, adotada neste trabalho, que sugere a possibilidade de um modelo misto/integrado que assume, além do tratamento estatístico, habilidades computacionais que permitam o tratamento simbólico dos dados, caracterizando-as como habilidades de domínio específico (GERVAIN & MEHLER, 2010; ENDRESS, NESPOR & MEHLER, 2009).

---

<sup>6</sup> Saffran & Thiessen (2007) dissertam sobre cinco aspectos da linguagem – Percepção da fala, Categorias da fala, Segmentação de palavra, Palavras e significados e sintaxe (do inglês “Speech perception, Speech Categories, Word segmentation, Words and meaning and Syntax”) - nos quais, tradicionalmente, o domínio de especificidade tem sido um foco explícito de investigação. Dos cinco aspectos, destacamos i) o fato de, enquanto estudos sobre desenvolvimento inicial ou sobre inatividade garantirem que a fala é altamente saliente em comparação a outros estímulos, há inúmeras evidências sugerindo que a **percepção da fala** é influenciada pelo mesmo mecanismo de aprendizagem responsável por processar outros tipos de estímulos; ii) a partir de um conjunto de evidências, que a trajetória de desenvolvimento subjacente a utilização e integração de múltiplas pistas na tarefa de **segmentação de palavras** dos bebês pode surgir a partir de processos de domínio geral e, por último que iii) no nível da **sintaxe**, apesar de haver evidências de que pelo menos algumas regularidades sintáticas podem ser aprendidas por mecanismos de domínio geral, a categorização através de informação distribucional é improvável de ser limitada a aprendizagem de línguas.

### **3.4- Proposta de integração**

De forma geral, mecanismos de aprendizagem estatística têm se mostrado robustos, de domínio geral, que se mantém ao longo da vida e como uma habilidade não específica para a espécie humana (GERVAIN & MEHLER, 2010). Uma questão central percorre essa proposta de integração: a aquisição de uma língua pelo bebê depende apenas da atuação de mecanismos estatísticos?

A literatura nos traz evidências de que o mecanismo estatístico pode operar sobre os sons da fala, como na entonação em pesquisa que sugere que recém nascidos foram capazes de aprender probabilidades transicionais e segmentar palavras estatisticamente a partir de uma sequência de sons contínuos (KUDO et al., 2006), e no estímulo visual como o estudo que sugere que bebês, aos 9 meses, foram capazes de aprender características visuais de ordem superior baseados na coerência estatística dos elementos entre das cenas, permitindo-lhes desenvolver uma representação eficiente para outras aprendizagens associativas (FISER & ASLIN, 2002). Além disso, tal mecanismo pode ser usado tanto por recém nascidos, quanto por bebês entre 8 e 13 meses (SAFFRAN, 2009) e adultos (PEÑA et al. 2002).

Mais que isso, macacos tamarinos, assim como humanos, mostraram evidências claras de discriminação entre sequências de sílabas que diferiam somente na frequência ou na probabilidade com que as quais ocorriam nos estímulos, sugerindo que seres humanos e primatas não-humanos possuem mecanismos capaz de computar aspectos particulares de ordem serial. (HAUSER et al. 2001) e estudos sugerem que roedores podem segmentar os fluxos usando a frequência de co-ocorrência<sup>7</sup> entre os itens, mostrando algum mecanismo básico de aprendizagem estatística generaliza nas espécies não-primatas (TORO & TROBALON, 2005). Entretanto, mesmo com tais evidências e conforme vimos (MARCUS et al., 1999; PEÑA et al., 2002; GERVAIN & MEHLER, 2010), adotar somente mecanismos estatísticos não seria suficiente para explicar a capacidade de tratamento do estímulo linguístico. Seria, dessa forma, plausível assumir a potencialidade de um mecanismo estatístico, não específico para a linguagem, mas indispensável associá-lo a uma habilidade de tratamento simbólico do estímulo.

---

<sup>7</sup> No estudo conduzido por Toro e Trobalon (2005), entretanto, roedores não diferenciaram entre os itens de teste quando o fluxo foi organizado por regularidades mais complexas que envolviam elementos não-adjacentes e regras abstratas de gramática.

Segundo proponentes dessa integração, essa proposta surge como uma proposta complementar na discussão sobre a aquisição da linguagem, focalizando funções perceptuais e de memória que restringem os estímulos apresentados no *input*. Assume-se que “mecanismos originalmente não linguísticos foram recrutados pela faculdade da linguagem e agora restringem os tipos de regularidades linguísticas que os aprendizes podem adquirir do *input*” (Endress et al. (2009), pág. 348, Tradução minha). Defendem, portanto, que, em termos evolutivos, os mecanismos que não eram linguísticos originalmente nas espécies humana e não humana agora, na espécie humana, restringem o tipo de regularidade que pode ser aprendida, ou seja, são mecanismos que são especializados em um domínio, mas não são específicos daquele domínio, pois foram provenientes de outro ou outros domínios. Assim, na evolução, essa faculdade\capacidade da linguagem teria se desenvolvido, recrutando mecanismos que estavam disponíveis no nosso aparato perceptual e cognitivo e que não eram específicos para a linguagem. Tais mecanismos continuam disponíveis em outros domínios, mas têm um uso especializado na capacidade linguística, tal qual hoje estaria disponível na espécie humana.

Ao analisar as atividades experimentais que envolvem o aprendizado de línguas artificiais, os autores sugerem que o sucesso dos aprendizes nessas tarefas está relacionado aos mecanismos computacionais especializados para uma função particular, os quais devem, provavelmente, derivar de mecanismos pré-existentes perceptuais ou de memória – Os POMPs.

Os POMPs – *perceptual or memory primitives* - (primitivos perceptuais ou de memória) são definidos como uma “caixa de ferramentas” (*a toolbox*), os quais poderiam complementar e restringir tanto relações baseadas em regras como relações estatísticas na aquisição e uso da linguagem e seriam capazes de explicar o aprendizado tanto em línguas artificiais como em línguas naturais. Segundo Endress et al. (2007), a mente pode ser equipada com essa “caixa de ferramenta” computacional de primitivos perceptuais especializados, que pode resolver alguns problemas computacionais. O termo "primitivos perceptuais", segundo os autores, foi escolhido para distinguir estas operações de operações primitivas mais gerais (como aquelas em um computador) e para destacar suas limitações intrínsecas. Os autores não afirmam que todas as operações mentais são “perceptuais”, e, ao contrário, sugerem que um “Gestalt especializado” e operações restrinidas, como as de detecção de repetições, podem ser encontrados em diferentes módulos e em diferentes níveis de abstração. Não apresentam, portanto, a intenção de tratar a “perceptualidade” de tais

operações, mas sim suas especializações e limitações, além do fato de serem recrutados sem muito esforço. Sobre os POMPs, afirmam:

In other words, although no other animal has language, certain linguistic computations might draw specialized, phylogenetically pre-existing mechanisms, and these mechanisms might determine what kinds of linguistic structures humans can learn.

(ENDRESS et al. 2009, p.348)

Segundo os autores, os mecanismos requeridos por alguns aspectos gramaticais das línguas são especializados, ou seja, podem ser usados no aprendizado de certas regularidades gramaticais, mas podem também restringir o tipo de regularidade a ser aprendida. O esquema abaixo representa os tipos de mecanismos supostos:

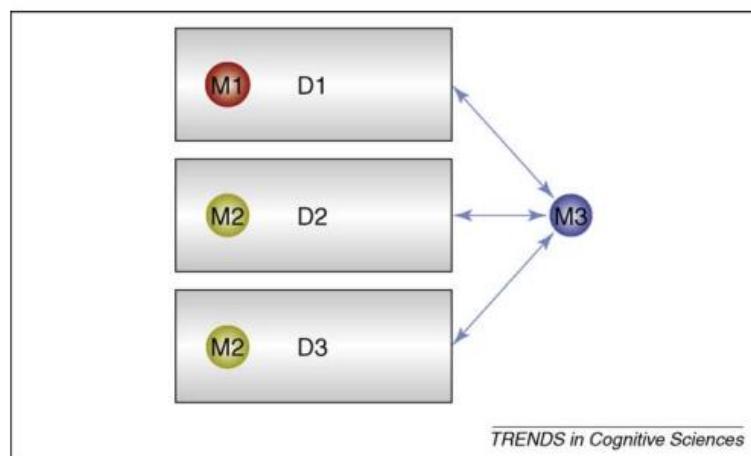


Figura 2 - Esquema extraído de Endress et al. 2009, pág. 349.

Segundo essa proposta, no esquema acima, os retângulos representariam os domínios, enquanto os círculos representariam os mecanismos. Assim, o mecanismo M1 (em vermelho) é de domínio específico<sup>8</sup> porque existe exclusivamente no domínio D1 e em nenhum outro

<sup>8</sup> Domain-specific mechanism: a specialized or non-specialized mechanism that exists exclusively in one particular domain. For instance, a mechanism to recognize colors despite illumination changes will be found only in vision, but not in other domains.

domínio. O mecanismo M3 (em azul) é de domínio geral<sup>9</sup> porque pode ser acessado por todos os domínios. O mecanismo M2 (amarelo) é domínio de fronteira<sup>10</sup> porque “cópias” (aspas dos autores) desse mecanismo existem em mais de um domínio (aqui em D1 e D3). Cada uma das cópias de M2 poderia operar somente em seu próprio domínio. Dessa forma, a disponibilidade dos primitivos especializados poderia determinar os tipos de regularidades que os aprendizes poderiam aprender mais facilmente.

Essa proposta de integração sugere, ao menos, dois mecanismos básicos que podem explicar a aquisição de uma língua: o primeiro diz respeito à **identificação das relações estabelecidas entre os elementos no *input***, enquanto o segundo usa como base o **início de sequências como pontos de ancoragem para extrair regularidades posicionais**.

A repetição das estruturas e as regularidades baseadas nas extremidades presentes em atividades que envolvem o aprendizado de gramáticas artificiais constituem dois mecanismos que facilitam o aprendizado em línguas artificiais e, por semelhança, a aquisição de línguas naturais.

A atividade desenvolvida por Marcus et al. (1999), descrita na seção anterior, é um clássico exemplo que sustenta a hipótese de um primitivo que é sensível à identificação de relações. Ao mostrar que bebês, aos 7 meses, são capazes de reconhecer uma dada estrutura e generalizá-la mesmo com o uso de novas sílabas durante o teste, os pesquisadores evidenciam que os bebês foram capazes de descobrir a relações entre essas variáveis. Dessa forma, as generalizações baseadas em repetição de padrões apresentadas nas atividades experimentais que envolvem língua artificial parecem ser computadas por um mecanismo especializado que seria “acionado” pela frequência da ocorrência de uma dada estrutura, subjacente aos dados na superfície, presente nos estímulos, fornecendo às mesmas um caráter importante no aprendizado de regras. Assim, segundo os autores, ainda que os POMPs sejam mais comuns nos casos de repetições de elementos adjacentes, o mesmo também parece ocorrer nos casos

<sup>9</sup> Domain-general mechanism: a specialized or non-specialized mechanism that can be accessed from different domains. Attentional processing might be an example.

<sup>10</sup> Domain-bound mechanisms: similar to domain-specific mechanisms, except that they are shared by at least two domains. For example, one can imagine distinct mechanisms for processing identity-relations in different domains such as vision, audition and language that are all independent of one another. That is, although these mechanisms might have very similar properties (they all process identity-relations), each of these mechanisms might operate in only one specific domain. For example, the identity-processing mechanism in vision might not process auditory identity-relations, and vice-versa. As domain-bound mechanisms are available in various domains, they are hard to distinguish from domain-general mechanisms (and, to our knowledge, have not been distinguished from them). The crucial difference, however, is that each of the mechanisms is implemented in such a way that it operates in only one particular domain.

nos quais as repetições se encontram um pouco mais distantes, como na relação de DNA dos estímulos das atividades experimentais desenvolvidas neste trabalho.

Além da POMP de repetição, a POMP baseada na posição dos itens na sequência também se faz presente nos experimentos sobre aprendizagem de gramáticas artificiais. Ainda citando o experimento de Marcus e colaboradores (1999), além de detectar a repetição de um item, os bebês, para terem sucesso na atividade, teriam que relacionar a repetição à sua posição estrutural. Endress e colaboradores (2005) mostraram que as generalizações baseadas na repetição de estrutura somente ocorrem quando as repetições estão presentes nas extremidades das sequências. Os resultados de Peña et al. (2002) também mostram que em enunciados, os participantes foram capazes de generalizar a regra quando as sílabas que faziam parte dessa gramática ocupavam as extremidades, diferentemente de quando as mesmas apareciam em posições intermediárias.

De acordo com essa proposta, entretanto, esses dois processos – detecção de repetições e memorização de sequência - seriam independentes, ou seja, deveríamos usar um POMP para detectar repetições e outro para detectar onde a sequência de repetições ocorre. Assim, a faculdade da linguagem estaria equipada com um conjunto de POMPs, os quais seriam, em parte, específicos para a linguagem e, em parte, “reciclados” de outros domínios – como demonstrado no esquema –, particularmente da percepção e da memória.

Dessa forma, ao investigarmos as etapas iniciais de aquisição lexical, buscamos discutir a abstração e generalização de dependências não-adjacentes e a categorização de itens da língua a partir da integração de mecanismos estatísticos e de habilidades computacionais/simbólicas para o tratamento das informações de natureza linguística. Pressupomos que o bebê não seja apenas sensível a características prosódicas da fala, mas também que seja dotado de mecanismos – como os POMPs – que lhe permitam analisá-las de modo a identificar propriedades morfonológicas e posicionais de itens lexicais. Uma caixa de ferramentas computacional de primitivos perceptuais pode também ajudar na integração dos modelos simbólicos e estatísticos na mente. Processos estatísticos e simbólicos podem contribuir com diferentes “ingredientes” para o aprendizado, visto que operações simbólicas – como os primitivos perceptuais – podem fornecer restrições representacionais sobre quais mecanismos de aprendizagem estatística podem operar. Todas essas pistas tomadas em conjunto, e quando disponíveis no estímulo, seriam responsáveis pelo processo inicial de aquisição da linguagem.

#### **4- ATIVIDADES EXPERIMENTAIS**

O presente capítulo dedica-se à apresentação das investigações experimentais realizadas. Na primeira seção, será apresentada a técnica experimental utilizada e a justificativa de seu uso. Na seção 4.2, a primeira atividade experimental desenvolvida – Relação entre DNA e Prosódia – será apresentada, bem como a elaboração e análise acústica dos estímulos utilizados, participantes, variáveis experimentais, hipóteses, metodologia e resultados. Essa atividade foi aplicada em dois grupos de participantes distintos: bebês canadenses (1A) e bebês brasileiros (1B). Dessa forma, os resultados serão apresentados separadamente. A seção 4.3 apresentará a segunda atividade experimental desenvolvida – DNA e categorização, a elaboração dos estímulos, participantes, variáveis experimentais, hipóteses, metodologia e resultados. A última seção apresentará uma discussão sobre como interpretar as reações das crianças aos estímulos, a partir do uso das técnicas experimentais “Olhar preferencial” e “Escuta preferencial”, ou seja, o que de fato avaliamos com a reação do bebê durante a atividade experimental.

O presente estudo faz uso de metodologia experimental. Ainda que gerem contextos artificiais de uso da língua para o estudo de aquisição lexical e processamento da linguagem, a metodologia experimental permite investigar o processo de aquisição lexical em bebês ainda em fase pré-verbal e controlar as variáveis que possam afetar o fenômeno investigado.

Duas atividades experimentais foram propostas. A primeira, dividida em duas amostras (bebês canadenses (1A) e bebês brasileiros (1B)) busca investigar se informações prosódicas alinhadas às informações de padrões de DNA podem facilitar a abstração e generalização de padrões de DNA. Essa atividade foi aplicada em duas etapas: na primeira etapa, a atividade foi desenvolvida em Montreal (Canadá) durante período de estágio de pesquisa (doutorado sanduíche) da autora desta tese, sob supervisão da professora Rushen Shi; na segunda etapa, a mesma atividade foi desenvolvida na UFJF, com bebês brasileiros. Ao eleger grupos distintos de participantes, buscamos confrontar os resultados, analisando o tipo de tarefa exigida para cada grupo de participantes. A segunda atividade desenvolvida propõe investigar se a partir do reconhecimento, abstração e generalização das DNAs, bebês brasileiros são capazes de categorizar os itens do léxico. Antes de detalharmos cada atividade, na próxima seção, apresentaremos a técnica experimental utilizada nas duas atividades desenvolvidas.

#### **4.1- Técnica Experimental – Olhar preferencial**

Em estudos com bebês (fase pré-verbal) que apresentam como objetivo principal verificar a percepção dos bebês a propriedades dos estímulos linguísticos, utilizam-se técnicas experimentais específicas, das quais se destacam a **Técnica de sucção não nutritiva**, **Técnica de escuta preferencial**, **Fixação preferencial do olhar**<sup>11</sup>, entre outras.

Por ser um procedimento adequado para avaliação de habilidades perceptuais em crianças com até os 18 meses, a técnica escolhida para a aplicação das atividades com bebês é uma variação da **Técnica de Escuta Preferencial** (*Head-Turn Preference Procedure* (HPP): KEMLER-NELSON ET AL., 1995).

Na Técnica de Escuta Preferencial, a criança é exposta a estímulos auditivos em eventos diferenciados em função de variáveis previamente selecionadas durante um período de tempo. Essa técnica parte do pressuposto de que o bebê reage se percebe diferença entre os estímulos apresentados, escutando por tempos diferentes os estímulos das duas condições distintas. Segundo Name (2012)

Esse paradigma explora uma relação de contingência entre estímulo e comportamento, de modo que a apresentação do estímulo é dependente do comportamento do bebê: se ele se interessa pelo que ouve, mantendo a cabeça virada para a direção de onde vem o som, o estímulo continua e ele escuta mais tempo; se desvia a cabeça, o som para e muda

Dividida em duas fases –familiarização e teste, é manipulada uma única variável em dois níveis. A fase de familiarização pode aparecer, ao menos, de três formas diferentes em

<sup>11</sup>**Sucção Não-Nutritiva:** adequada para bebês recém-nascidos e bebês de até 4 meses, tem como objetivo geral avaliar as habilidades discriminatórias disponíveis em face de estímulos linguísticos em que se manipulam propriedades acústicas. É dada ao bebê uma chupeta ligada a um computador. As taxas de sucção das fases de habituação e teste são comparadas e, se houver um aumento estatisticamente significativo da frequência de sucção a partir da mudança de estímulo ocorrida no grupo experimental, esse aumento sugere que há discriminação, por parte do bebê, das diferentes realizações da variável independente manipulada. **Fixação preferencial do olhar:** adequado para crianças de 4 meses a 4 anos de idade , explora o fato de o bebê, desde muito cedo, expressar a percepção de algo novo, fixando o olhar por mais tempo no objeto ou evento que difere em algum aspecto daqueles a que se encontra habituado.Há duas versões: a intramodal em que se lida com uma única modalidade perceptual, e a intermodal – usada nos estudos de aquisição da linguagem – que explora a capacidade de o bebê relacionar estímulos captados por diferentes modalidades perceptuais e tomá-los como vinculados a um único evento. Para mais detalhamento das técnicas aqui citadas, ver Name& Corrêa (2006).

relação ao objetivo proposto. A primeira tem como objetivo familiarizar o bebê com o *modus operandi* da atividade, ou seja, fazê-lo perceber que o estímulo sonoro sai apenas de um dos lados e apenas enquanto ele estiver olhando para a direção do som ou ainda acostumar o bebê a um dado estímulo linguístico de forma a reconhecê-lo em ensaios durante o teste. O segundo tipo tem como objetivo habituar o bebê previamente a um dado item do léxico (ou conjunto de itens, ou a uma dada estrutura) para, em seguida, identificá-lo nos enunciados. O terceiro tipo, utilizado nas atividades experimentais desenvolvidas nesta tese, tem como objetivo expor a criança a um conjunto de estímulos com alguma regularidade (concordância morfossintática, p.ex.), durante um determinado tempo, sensibilizando-a aos estímulos. Name (2012) sugere que os três tipos de familiarização, podem ser denominados, respectivamente de “aquecimento”, “habituação” e “sensibilização”.

Cabe ressaltar que os dois primeiros tipos de familiarização são controlados pela criança, o que significa que se ela se interessa pelo que está ouvindo, olha na direção do som; se não há interesse, desvia o olhar/a cabeça e o som para. Diferentemente, no terceiro tipo de familiarização - sensibilização, a apresentação do estímulo tem duração fixa (comumente, dois minutos) e não depende da atenção do bebê. Nesse tipo de familiarização, o bebê não precisa olhar na direção do som durante os dois minutos de familiarização, sendo necessário que ele apenas esteja “disponível” ao estímulo sonoro.

Na fase de teste, apresentam-se à criança dois tipos de estímulos distintos, coerentes ou não com o que foi anteriormente apresentado. Os estímulos são apresentados em áudio, por meio de dois alto-falantes situados à esquerda e à direita da criança. O experimentador cronometra o tempo de escuta da criança - o tempo que ela se volta para o lado de onde vem o som do alto-falante (a variável dependente: o tempo de escuta , i.e., a medida comportamental do sujeito).

A variação dessa técnica, usada em diversos laboratórios psicolinguísticos e que foi usada nesta pesquisa, é a técnica do **Olhar Preferencial** (*Preferential Looking Procedure*). O procedimento é basicamente o mesmo. A principal mudança é o uso de apenas um alto-falante, centralizado junto a uma tela maior, o que permite a simplificação do *software* usado (*Habit*<sup>12</sup>), sem perder a confiabilidade dos resultados. Uma câmera escondida filma o olhar da criança e o experimentador mede o tempo de fixação do olhar que corresponde ao tempo de escuta do estímulo acústico. A medida pode ser definida como o olhar da criança ao estímulo visual apresentado. O experimentador pode aferir, com maior precisão e controle, o

---

<sup>12</sup>(Cohen, Atkinson & Chaput, 2000). Trata-se de um software gratuito que exige a plataforma MAC.

tempo de olhar do bebê, que é considerado o tempo de escuta do estímulo apresentado ou, ao menos, o tempo que o bebê presta mais atenção àquele estímulo. Além da familiarização e do teste, o Olhar Preferencial conta com uma fase de preteste e uma de posteste.

A atividade começa com uma imagem na tela diante do bebê, acompanhada por um som não linguístico, para chamar a atenção do bebê e fazê-lo olhar para frente (evento chamado de *attention-getter*). Capturada a atenção da criança, inicia-se o preteste, em que é apresentada junto com uma pequena história, uma imagem que também será apresentada durante o teste (não muito atrativa para as crianças, visto que a imagem apresentada não pode ser mais atrativa que o estímulo linguístico). Em seguida, novamente a tela de *attention-getter* aparece e a familiarização começa, com os estímulos linguísticos acompanhados da mesma imagem apresentada no preteste. Após a fase teste, o posteste, no qual é apresentada uma pequena história semelhante a do pré teste e com a mesma imagem das fases anteriores) visa a assegurar o interesse da criança até o final da atividade. Entre cada fase e entre cada ensaio da fase de teste é apresentada a tela *attention-getter*, de modo a garantir a atenção do bebê no início de cada fase e ensaio.

A confiabilidade dos resultados se torna possível porque o *Habit* é programado para fazer com que o estímulo auditivo seja interrompido caso o participante desvie o olhar por mais de 2 segundos consecutivos. Esse mecanismo faz com que o *attention-getter* (tela que tem o objetivo de chamar a atenção do bebê – centralizar seu olhar) seja acionado e, após a retomada da atenção por parte do bebê, o próximo estímulo seja apresentado.

Torna-se bastante importante ressaltar que não há variação da imagem apresentada, diferentemente dos estímulos acústicos, que variam de acordo com as condições de teste. Além disso, do lado de fora da cabine, o experimentador não sabe que tipo de estímulo a criança está escutando, de modo que não há qualquer interferência de sua parte, ainda que involuntária, na medição do tempo de fixação do olhar. Adiante, serão apresentadas as descrições de todas as etapas e os procedimentos que envolveram a atividade experimental. As duas atividades experimentais desenvolvidas foram compostas de quatro fases – habituação (pré-teste), familiarização, teste e pósteste – que serão descritas nas seções seguintes.

#### **4.2- Atividade Experimental 1 - Relação entre DNA e prosódia (1A)**

O objetivo desta atividade experimental é investigar se informações prosódicas alinhadas às informações de padrões de DNA podem facilitar a abstração e generalização de padrões de DNA. Em estudo realizado anteriormente (TEIXEIRA, 2012; TEIXEIRA; NAME, 2012), verificamos que bebês brasileiros, com idade média de 11 meses, a partir de curta exposição a uma língua pseudonatural com padrões de combinações entre D e N diferentes de sua língua materna, foram capazes de reconhecer padrões de relações entre D e N e, ainda, de abstraí-los de modo a identificá-los em novos enunciados no âmbito do DP.

Vimos no capítulo 2 desta tese que DNAs, constituídas em sua maioria de elementos funcionais, se apresentam frequentemente em fronteiras de domínios prosódicos, especialmente em fronteiras de sintagmas fonológicos e que esse constituinte prosódico, além de ser um dos mais frequentes no PB, é importante para a interface sintaxe-prosódia. Buscamos, dessa forma, avaliar em que medida o alinhamento de DNAs às fronteiras de sintagmas fonológicos facilita ou potencializa a abstração e generalização (representação) das DNAs. A fim de avaliar se aos 11 meses a criança é capaz de identificar uma propriedade comum às línguas naturais através da percepção de DNA nos estímulos linguísticos a que tem acesso durante um breve espaço de tempo, buscamos (1) criar condições prosódicas semelhantes às existentes no PB para alocar as DNAs, e (2) investigar qual condição representaria um ambiente de maior “aprendibilidade” dos padrões de DNAs.

Como foi apresentado anteriormente, o diferencial dessa atividade em relação às já existentes na literatura é que buscamos preservar propriedades inerentes às línguas naturais, como os itens funcionais (presentes nas DNAs) e o contorno prosódico – diferentemente das atividades que utilizam línguas totalmente artificiais. Sabe-se que no PB, há estruturas que contêm relações de DNAs e que são muito parecidas segmentalmente, mas são distintas em relação ao contorno prosódico. Estruturas como Det +N+ Adj e Det + N +V, como nos exemplos apresentados abaixo, apresentando o mesmo determinante (Essa), o mesmo Nome (menina) e elementos pospostos ao nome que poderiam pertencer a mesma classe gramatical.

(18) Essa menina gentil

(19) Essa menina dormiu

Como falantes do PB, sabemos que o Det “essa” foi utilizado por concordar em gênero e número com o nome “menina”. Mais que isso, sabemos que na primeira sentença a palavra “gentil” pertence à classe dos adjetivos e “dormiu” à classe dos verbos. Entretanto, tal distinção aparece somente no nível estrutural, visto que um falante que não conhece a língua e seus morfemas (apenas lendo as sentenças), dificilmente conseguiria distinguir os últimos elementos apresentados nos exemplos acima (adjetivo e verbo). Entretanto, ao ouvirmos as sentenças, também como falantes nativos do PB, sabemos que as duas sentenças apresentam contornos prosódicos distintos, visto que a primeira trata-se de um DP e a segunda de VP e, muito provavelmente, conseguiríamos discriminá-las aos escutá-las. Enquanto a primeira apresenta uma fronteira de sintagma fonológico após o adjetivo, a segunda apresenta a fronteira de sintagma fonológico entre N e V, caracterizando uma prosódia distinta entre as duas sentenças.

(20) [Essa menina gentil]  $\emptyset$

(21) [Essa menina]  $\emptyset$  [dormiu]  $\emptyset$

Dessa forma, notamos que, enquanto na sentença (20) a relação de DNA entre D e o morfema de gênero (**Essa menina**) encontra-se distante da fronteira de sintagma fonológico, na sentença (21) a relação de DNA entre D e morfema de gênero (**Essa menina**) encontra-se alinhada à fronteira. Ora, se bebês brasileiros ao 11 meses de idade foram capazes de reconhecer padrões de relações entre D e N e, ainda, de abstraí-los de modo a identificá-los em novos enunciados no âmbito do DP a partir de curta exposição a uma língua pseudonatural com padrões de combinações entre D e N diferentes de sua língua materna (Teixeira, 2012) e se sintagmas fonológicos representam ambiente produtivo para a interface sintaxe-prosódia, o conjunto dessas informações/pistas, ou seja, o alinhamento de DNA às fronteiras de sintagmas fonológicos, deverá facilitar a abstração e generalização dos padrões das DNAs.

Cabe ressaltar que no DP (Det+N+Adj) há uma fronteira de palavra prosódica depois de N. Entretanto, neste trabalho, não estamos investigando a presença de qualquer fronteira, nem mesmo a dualidade entre presença de fronteira x ausência de fronteira, mas sim a dualidade entre presença de fronteira de sintagma fonológico entre N e V x ausência de fronteira de sintagma fonológico entre N e ADJ, ressaltando que tanto no DP quanto no VP há o mesmo tipo de DNAs entre Det e terminação do N. Contudo, no VP a DNA encontra-se alinhada à fronteira de sintagma fonológico.

A partir dessas observações de dados do PB e da hipótese gerada, os estímulos foram criados, conforme será apresentado na seção a seguir.

#### **4.2.1- Elaboração dos estímulos**

Para a construção dos estímulos, foram criadas duas condições experimentais: Condição Adjetivo e Condição Verbo. As duas condições possuem o mesmo número de elementos (três), os mesmos “elementos” superficialmente (Det+N+Adj ou Det+N+V) e o mesmo padrão (combinação) de DNA (entre o determinante e a terminação no N). Dessa forma, a única diferença entre as duas condições é o contorno prosódico característico de um DP ou de VP.

Nesta atividade, foram utilizados 4 (quatro) determinantes do PB (2 masculinos e 2 femininos): **Nossa, Essa, Seu e Meu**; 6 (seis) pseudopalavras dissílabas terminadas em **-e** e 6 (seis) pseudopalavras dissílabas **-o**.

Sabe-se que os determinantes no PB podem ser monossílabos, dissílabos ou trissílabos. Decidimos, nesta atividade, não utilizar determinantes monossílabos simples (como “a” e “o”) por não apresentarem saliência fonológica, assim como não utilizar determinantes trissílabos (como “aquele” ou “aquela”) para não aumentarmos a janela de processamento entre o determinante e a terminação de nome, visto que não era esse o nosso objeto de investigação. Os determinantes femininos, em PB, sempre terminam em **-a**. Como não queríamos dois determinantes com fonemas parecidos (como os nasais /m/ e /n/, de “minha” e “nossa”), optamos por utilizar dois determinantes femininos dissílabos e distintos: “nossa” e “essa”. Como as crianças são expostas aos quatro determinantes, o determinante masculino “nosso” foi descartado, por apresentar semelhança à forma feminina utilizada “nossa”. Optamos por utilizar, portanto, dois determinantes monossílabos masculinos “meu” e “seu”.

Sabemos que em PB a terminação padrão dos nomes femininos é **-a**, enquanto que a terminação **-o** é padrão em nomes masculinos. Entretanto, no PB, temos palavras com a terminação **-o** que são femininas e em **-e** que podem ser masculinas ou femininas, por isso escolhemos essas terminações. Para a criação dos pseudoadjetivos e pseudoverbhos, buscamos uma terminação que fosse possível no PB e que apresentasse a mesma terminação, em termos fonéticos, visto que iríamos utilizá-los nas duas condições com o objetivo de que somente a prosódia fosse um fator diferenciador entre elas. Dessa forma, escolhemos a terminação [iw]

do PB (grafada como *-il* para adjetivos – “gentil”, “febril” etc. – e como *-iu* para verbos – como os verbos da terceira conjugação: “dormiu”, “partiu”, “saiu”...), totalizando 23 pseudoverbos/pseudoadjetivos. Para a criação dos pseudonomes, assim como dos pseudoadjetivos e pseudoverbos, buscamos manter o padrão silábico do PB (CV).

Foram criados dois padrões/combinações de DNA entre Det e terminação do N nas duas condições apresentadas a seguir:

Condição Adjetivo: [Nossa/Essa + N-o + Adj-[iw]]<sub>φ</sub> e [Meu/Seu + N-e + Adj-[iw]]<sub>φ</sub>

Condição Verbo: [Nossa/Essa + N-o]<sub>φ</sub> + [V-[iw]] e [Nossa/essa + N-e]<sub>φ</sub> + [V-[iw]]

Para evitarmos alguma preferência por um padrão em detrimento de outro, criamos duas combinações de DNA em cada condição, controlando para que uma possível preferência por um determinado padrão não interferisse em nossos resultados. Nas tabelas abaixo, apresentamos as duas condições e as quatro combinações utilizadas da atividade experimental:

<b>Condição Verbo</b>			
<b>[Det + N]<sub>φ</sub> [V]</b>			
<b>Grupo</b>	<b>Det</b>	<b>Pseudonome</b>	<b>Pseudoverbo</b>
<b>A</b>	Nossa/Essa	x-o vato	y- [iw] muniu
	Meu/Seu	x-e dale	y- [iw] biliu
<b>B</b>	Nossa/Essa	x-e lepe	y- [iw] pigiu
	Meu/Seu	x- o pifo	y- [iw] cutiu

Tabela 1 - Combinação entre Determinantes e terminação dos N na condição Verbo

<b>Condição Adjetivo</b>			
<b>[Det + N + Adj]<sub>Φ</sub></b>			
<b>Grupo</b>	<b>Det</b>	<b>Pseudonome</b>	<b>Pseudoadjetivo</b>
<b>A</b>	Nossa/Essa	x-o vato	y- [iw] munil
	Meu/Seu	x-e dale	y- [iw] bilil
<b>B</b>	Nossa/Essa	x-e lepe	y- [iw] pigil
	Meu/Seu	x- o pifo	y- [iw] cutil

Tabela 2 - Combinação entre Determinantes e terminação dos N na condição Adjetivo

As listas completas dos estímulos nas duas condições podem ser visualizadas no anexo 1.

#### 4.2.2 – Análise acústica dos estímulos

Como o presente trabalho objetiva investigar se a presença de fronteira de sintagma fonológico alinhada à DNA (entre Det e terminação de N) é um fator facilitador da percepção, abstração e generalização do padrão da DNA, elaboramos condições experimentais que pudessem suportar esse ambiente. Dessa forma, como dito anteriormente, duas condições experimentais foram criadas – Condição Adjetivo e Condição Verbo.

Como vimos, a única diferença entre as duas condições é a presença da fronteira de sintagma fonológico alinhada à DNA na condição Verbo e, consequentemente, um contorno prosódico característico desta condição. Todas as sentenças foram gravadas por uma falante nativa do PB e a leitura das frases foi realizada do modo mais natural possível. As gravações foram capturadas através de um gravador de voz digital de alta qualidade (SONY PCM-D50) e realizadas no Laboratório de Atividades Experimentais do Groupe de Recherche sur le Langage da UQAM/ Montreal , ambiente com isolamento acústico, que eliminou a interferência de ruídos externos. Utilizando a voz de uma só participante, objetivamos a uniformidade entre as sentenças, para que fosse possível comparar as propriedades prosódicas que sinalizam uma fronteira de sintagma fonológico e a garantia de que todos os bebês testados fossem expostos aos mesmos estímulos linguísticos.

As gravações obtidas para o desenvolvimento da atividade experimental foram analisadas acusticamente, considerando-se os seguintes parâmetros suprasegmentais: (i) a

**duração** (das pausas – da condição Verbo - e das sílabas pré e pós-fronteira), em milésimos de segundo; (ii) a **intensidade** (das sílabas pré e pós-fronteira), bem como seus valores máximos e mínimos. Esses valores foram obtidos através do *software* PRAAT e determinados com base nas informações auditivas e na interpretação do espectrograma e da forma da onda sonora dos estímulos.

Dessa forma, todos os parâmetros descritos acima foram analisados nos estímulos das seguintes fases do experimento: familiarização (totalizando 4 listas: 2 grupos em cada condição) e fase teste (das duas condições). Para melhor compreensão, apresentaremos, a seguir, um exemplo dos estímulos utilizados, em que podemos visualizar, em destaque, os segmentos analisados:

Condição Adjetivo:  $\phi$  [Essa dabo munil]  $\phi$

Condição Verbo:  $\phi$  [Essa dabo]  $\phi$  [muniu]  $\phi$

Abaixo, podemos visualizar - a título de exemplificação – a diferença entre as duas sentenças no espectrograma. Observamos, neste primeiro momento, a presença da pausa sinalizando o sintagma de fronteira fonológica entre N e V na condição Verbo.

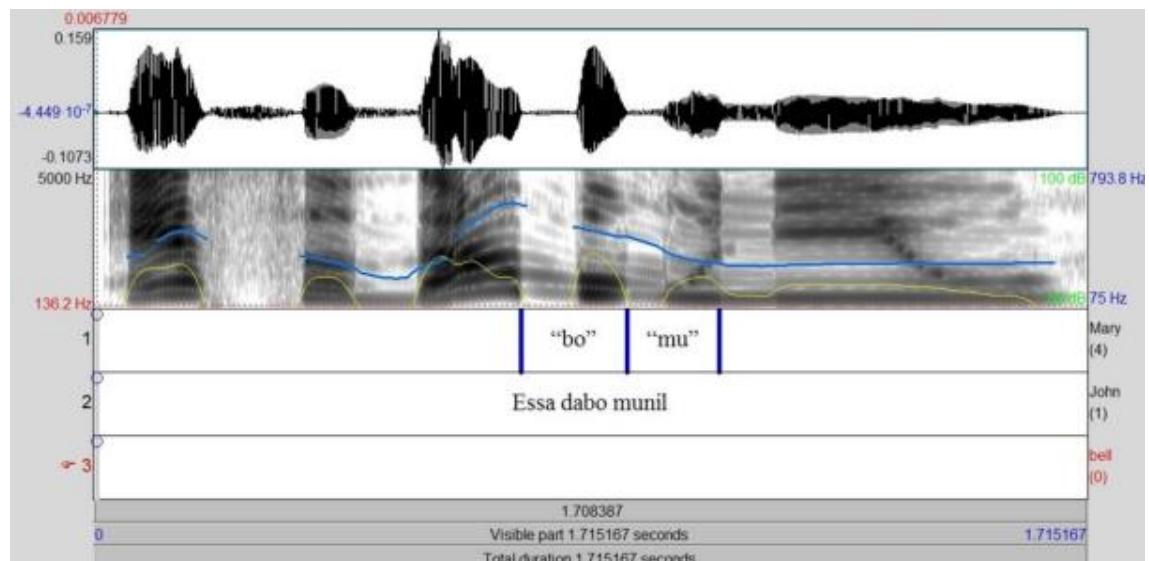


Figura 3 – Espectrograma do estímulo “Essa dabo munil” na condição Adjetivo

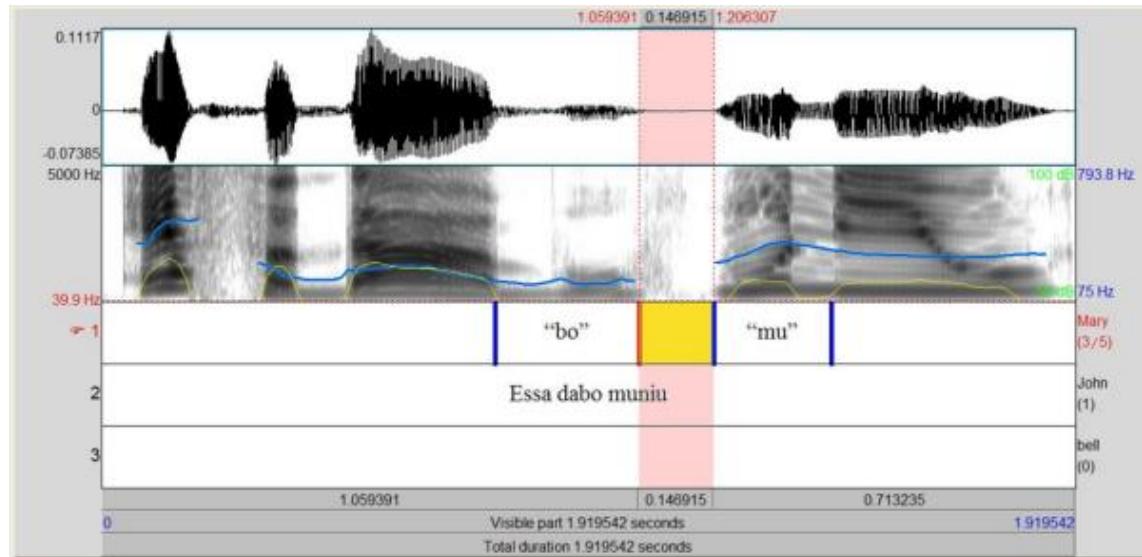


Figura 4 - Espectograma do estímulo “Essa dabo muniu” na condição Verbo

O primeiro segmento analisado foi a pausa, característica de fronteira de sintagma fonológico. O parâmetro analisado dessa pausa foi a duração. Identificamos a média de 0,09 s nos estímulos da fase de familiarização (dos grupos 1 e 2) e 0,11s na fase teste, ambos da condição Verbo. Na condição Adjetivo, não foram encontradas pausas significativas, conforme prevíamos, devido à ausência de fronteira de sintagma fonológico. Além da pausa, decidimos realizar a análise dos segmentos que imediatamente antecediam ou precediam a fronteira prosódica.

O segundo segmento analisado foi a última sílaba do nome em ambas as condições. Observamos, nesse segmento, maior duração do segmento na condição Verbo e maior intensidade na condição Adjetivo (ainda que pequena). Em relação aos parâmetros intensidade mínima e máxima, o segmento da condição Adjetivo apresentou valores maiores. As médias dos valores de cada parâmetro, neste segmento, podem ser observadas na tabela abaixo:

Última sílaba do Nome				
	Duração	Intensidade	Intensidade mínima	Intensidade máxima
Condição Adj Grupo A	0,1552	<b>54,01</b>	<b>42,04</b>	<b>58,44</b>
Condição Adj Grupo B	0,1632	<b>51,33</b>	<b>40,37</b>	<b>55,78</b>
Condição Adj Teste	0,1726	<b>54,52</b>	<b>43,49</b>	<b>58,29</b>
Média geral Condição Adj	0,1636	<b>53,28</b>	<b>41,96</b>	<b>57,50</b>
Condição V Grupo A	<b>0,2572</b>	51,00	35,48	54,84
Condição V Grupo B	<b>0,2290</b>	50,93	38,49	55,28
Condição V Teste	<b>0,2673</b>	49,31	40,50	53,63
Média geral Condição V	<b>0,2511</b>	50,41	38,15	54,58

Tabela 3 – Média dos valores de Duração, Intensidade, Intensidade Mínima e Intensidade Máxima da última sílaba do Nome

O terceiro segmento analisado foi a primeira sílaba do Adjetivo/Verbo. Observamos, nesse segmento, maior duração e maior intensidade na condição Verbo. Em relação ao parâmetro intensidade mínima, a condição Adjetivo apresentou maiores médias e, em relação ao parâmetro intensidade máxima, a condição Verbo apresentou maiores médias. As médias dos parâmetros podem ser encontradas na tabela abaixo:

Primeira sílaba do Adjetivo/Verbo				
	Duração	Intensidade	Intensidade mínima	Intensidade máxima
Condição Adj Grupo A	0,1857	59,99	<b>41,65</b>	63,84
Condição Adj Grupo B	0,1832	59,29	<b>41,70</b>	63,04
Condição Adj Teste	0,1521	59,50	<b>51,91</b>	61,81
Média geral Condição Adj	0,1736	59,59	<b>45,08</b>	62,89
Condição V Grupo A	<b>0,2367</b>	<b>60,49</b>	38,01	<b>65,44</b>
Condição V Grupo B	<b>0,2293</b>	<b>59,33</b>	39,77	<b>64,05</b>
Condição V Teste	<b>0,2673</b>	<b>63,92</b>	44,37	<b>68,05</b>
Média geral Condição V	<b>0,2444</b>	<b>61,24</b>	40,71	<b>65,84</b>

Tabela 4 – Média dos valores de Duração, Intensidade, Intensidade Mínima e Intensidade Máxima da primeira sílaba do Adjetivo/Verbo

As análises apresentadas acima apontam, de fato, para a existência de fronteira de sintagma fonológico entre N e V. Como veremos, no gráfico a seguir, na condição Verbo há um expressivo alongamento de fronteira em relação à condição Adj. Isso significa dizer que em sintagmas como [Essa **dabo** **munil**] φ (Condição Adjetivo) e [Essa **dabo**] φ [**muniu**] φ (condição Verbo), houve um alongamento dos segmentos selecionados na condição Verbo devido à presença da fronteira fonológica entre N e V.

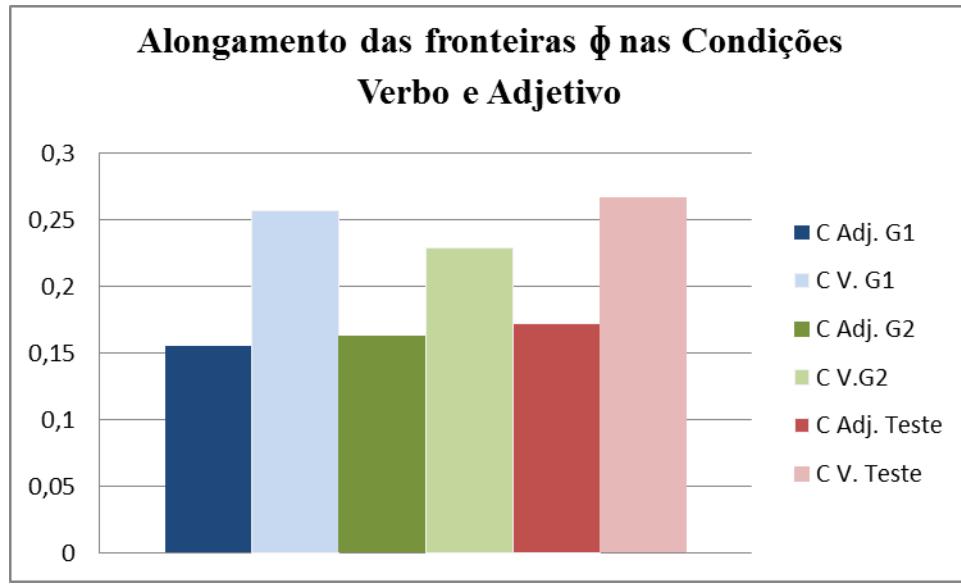


Gráfico 1 – Alongamento das fronteiras de sintagma fonológico nas Condições Verbo e Adjetivo

#### 4.2.3- Variáveis experimentais:

##### Variável independente:

Alinhamento dos elementos de DNAs a bordas de fronteiras prosódicas, em dois níveis (alinhado x não-alinhado)

**Variável dependente:** tempo de escuta (olhar).

##### Condições Experimentais:

- Condição Verbo: alinhamento dos elementos de DNAs a bordas de fronteiras de sintagma fonológico

$\phi [Nossa/Essa+N-o] \phi + [V-[iw]] \phi$  e  $\phi [Nossa/essa+N-e] \phi+[V-[iw]]$

- Condição Adjetivo: não alinhamento dos elementos de DNAs a bordas de fronteiras de sintagma fonológico.

$\phi [Nossa/Essa+N-o+Adj-[iw]] \phi e \phi [Meu/Seu+N-e+Adj-[iw]]\phi$

### Hipótese:

Elementos relacionados entre si em uma configuração de DNAs alinhados a bordas de fronteiras de sintagmas fonológicos são mais facilmente percebidos, abstraídos e generalizados por bebês aos 11 meses de idade do que quando ao menos um desses elementos não se encontra em borda de fronteira prosódica.

### Previsão:

Se DNAs alinhadas a bordas de fronteiras de sintagma fonológico são mais facilmente percebidas, abstraídas e generalizadas do que quando ao menos um desses elementos não se encontra em borda de fronteiras de sintagma fonológico, bebês aos 11 meses de idade serão capazes de abstrair o padrão e generalizá-lo em novos estímulos na condição em que há alinhamento dos elementos de DNAs a bordas de fronteiras de sintagma fonológico (Condição Verbo) do que na condição em que os elementos de DNAs não estão alinhados a bordas de fronteiras de sintagma fonológico (Condição Adjetivo).

#### 4.2.4- Método

##### Participantes

Participaram da atividade experimental 41 bebês canadenses<sup>13</sup> com idade média de 11 meses (11meses e 0 dias a 11 meses e 29 dias). Entretanto, 9 bebês foram eliminados por inquietação (4), choro (3) e problemas técnicos durante a sessão (2). Assim, os resultados encontrados são referentes a 32 bebês, dentre os quais 15 são do sexo feminino e 17 são do sexo masculino. Os 32 sujeitos foram divididos em dois grupos. Dessa forma, 16 crianças constituem o Grupo Adjetivo – Condição Adjetivo (8 do sexo feminino e 8 do sexo

---

<sup>13</sup> Bebês canadenses da cidade de Montreal.

masculino) e 16 constituem o Grupo Verbo– Condição Verbo (9 do sexo feminino e 7 do sexo masculino). Em cada grupo, 8 participantes constituem o grupo A e 8 participantes constituem o grupo B. A diferença entre grupo A e grupo B era somente a combinação entre Det e terminação no nome. Dessa forma, o grupo B apresentava padrão de combinação inverso do grupo A. Em cada subgrupo A ou B, existiu uma outra subdivisão em relação a ordem de apresentação dos estímulos testes. Assim, quatro crianças do grupo A, em ambas as condições, foram testadas com os estímulos testes na seguinte ordem “1, 2, 3, 4...”, enquanto que quatro crianças do grupo B, em ambas as condições, foram testadas com a seguinte ordem dos estímulos testes “2, 1, 4, 3...”. Essas subdivisões foram realizadas a fim de se controlar qualquer preferência em relação a um ou outro padrão/ordem.

Os participantes foram recrutados pelos pesquisadores do Groupe de Recherche sur le Langage da UQAM/ Montreal e receberam uma retribuição simbólica de CAN \$10 pela participação na atividade. Os responsáveis foram contatados via telefone e informados que a atividade é aprovada pelo Comitê de Ética da instituição, e foram orientados a assinar um Termo de Consentimento livre e Esclarecido da própria instituição, caso concordassem em participar da pesquisa.

### **Material e fases da atividade experimental**

Os equipamentos utilizados para a realização da atividade experimental foram: fones de ouvido, aparelho reproduutor de som MP3, monitor 42” MAC, filmadora Sony, caixa preta (“esconderijo” da câmera), Computador MAC APPLE G5, televisor, monitor Samsung 15” e teclado.

O experimento é constituído de quatro fases consecutivas que serão, abaixo, detalhadas. Essas fases são necessárias para permitir um maior controle dos fatores que poderiam influenciar o resultado, como por exemplo, garantir a “atenção” do bebê.

#### **1- Habituação**

A fase de habituação, também conhecida como preteste, consiste em um “aquecimento”. Nesta fase é apresentado auditivamente para a criança um estímulo contínuo de uma pequena história produzida com características de FDC (fala dirigida à criança). O

estímulo auditivo dura, em média, 8 segundos e é apresentado simultaneamente a um estímulo visual (o mesmo que será utilizado nas fases de familiarização e teste):

*“Olá! Que bom que você veio. Você quer brincar comigo? Vai ser super divertido!”*

## 2- Familiarização

A fase de familiarização, do tipo sensibilização, consiste em um áudio contínuo de 2 minutos constituído de sequências de DPs ou VPs<sup>14</sup> (a depender da condição escolhida) formados por determinantes, pseudonomes e pseudoadjetivos/pseudoverbos.

Para a fase de familiarização, foram criadas duas listas de cada condição, cada uma contendo 24 DPs ou VPs cada (a variar com a prosódia característica da condição) gravada por uma falante nativa do PB. A diferença entre as duas listas da condição DP ou da condição VP era apenas o padrão de combinação entre Det e terminação do Nome. Na construção da lista, foi introduzida uma pausa silenciosa de 1 segundo entre cada DP ou VP em ambas as condições. Durante a familiarização os bebês ouviram a um áudio que continha a lista duas vezes, totalizando 48 DPs ou VPs com duração total em cada lista de 129,40s, 128,15s, 138,88s e 135,91s de cada condição, respectivamente, para as condições Adjetivo A e B e Verbo A e B. Na fase de familiarização, foram utilizados 6 pseudonomes terminados em *-o* e 6 pseudonomes terminados em *-e*, além de 24 pseudoadjetivos/pseudoverbos distintos terminados em *-il/-iu*.

Durante esta fase, não nos era importante o rastreamento do comportamento visual da criança, pois o importante era a exposição aos estímulos da pseudolíngua. Assim, durante dois minutos, a criança estava diante dos estímulos contendo as relações de DNA dos DPs ou VPs . Cada uma das 12 palavras aparecia 2 vezes, cada vez precedida por um determinante diferente. Assim, “vato”, no grupo A, para ambas as condições, apareceria tanto antecedido pelos determinantes “nossa” quanto por “essa”, enquanto que, no grupo B, também em ambas as condições, apareceria tanto antecedido por ‘meu’ quanto “seu”.

## 3- Teste:

---

<sup>14</sup> DP do inglês Determiner Phrase e VP do inglês Verbal Phrase. Termos equivalentes a sintagma nominal e sintagma verbal, respectivamente, do português.

Para os ensaios da fase de teste, foram utilizados os 4 determinantes (nossa, essa, meu e seu) da familiarização e 4 novos pseudónomes (bifo, mape, rufe e leto) que não apareceram anteriormente na fase de familiarização (2 de cada terminação). Para preencher a posição de Adj ou V optamos por utilizar um pseudoadjetivo/pseudoverbo diferente dos apresentados na fase de familiarização, mas o mesmo Adj/V (duvil –Adj e duviu- V) em todos os ensaios, pois o foco era o reconhecimento ou não do padrão de DNA, portanto, variaram-se os itens envolvidos nas DNAs mas manteve-se o mesmo pseudoV ou pseudo ADJ nos diferentes trials. Dessa forma, foram criados 12 ensaios, cada um com 8 DPs ou VPs. Os ensaios dos dois grupos de cada condição eram idênticos, visto que o que determinaria se um ensaio seria grammatical ou agrammatical seria o padrão anteriormente familiarizado. Uma pausa silenciosa de 50 milisegundos foi inserida no início de cada ensaio e de 1 segundo entre os DP ou VP, assim como na familiarização. Cada ensaio apresentava 21,61s, 20,72s, 22,91s e 23,35s, respectivamente, para as condições Adj grupos 1 e 2 e V grupos 1 e 2. Para que possuíssemos uma mesma duração em cada condição (exigida pelo software utilizado), foi necessário fazer algumas reduções aleatórias de silêncio entre os DP e VP, totalizando a duração de 20,57s e 22,76s, respectivamente, para as condições Adj e V.

Duas condições experimentais foram criadas: CONDIÇÃO GRAMATICAL e CONDIÇÃO AGRAMATICAL. Na condição grammatical, a combinação entre “Det + terminação do novo Nome” possuía o mesmo padrão apresentado anteriormente na fase de familiarização, enquanto que na condição agrammatical o padrão não era congruente ao apresentado na fase de familiarização. É importante ressaltar que os ensaios na fase teste foram os mesmos para os dois grupos em cada condição. Dessa forma, o ensaio 1, por exemplo, é grammatical para o Grupo A e agrammatical para o Grupo B; o ensaio 2 é agrammatical para o Grupo A e grammatical para o Grupo B; o ensaio 3 é grammatical para o Grupo A e agrammatical para o Grupo B e assim por diante.

Dessa forma, no Grupo A das duas condições (N ou V), na condição grammatical, *nossa* e *essa* eram combinados com pseudónomes com terminação *-o* e *meu* e *seu* eram combinados com novos pseudónomes com terminação *-e*, enquanto que, na condição agrammatical, *nossa* e *essa* eram combinados com novos pseudónomes com terminação *-e* e *meu* e *seu* eram combinados com novos pseudónomes com terminação *-o*. De forma análoga, no Grupo B das duas condições, na condição grammatical, *nossa* e *essa* eram combinados com novos pseudónomes com terminação *-e* e *meu* e *seu* eram combinados com novos pseudónomes com terminação *-o*, enquanto que na fase agrammatical, *nossa* e *essa* eram

combinados com novos pseudónomes com terminação *-o* e *meu* e *seu* eram combinados com novos pseudónomes com terminações *-e*, como podemos visualizar na tabela abaixo:

		<b>Condição Verbo</b>		<b>Condição Adjetivo</b>	
		([Det+N] Φ [V])		([Det+N+Adj] Φ)	
Grupo	Det	Pseudonome	Pseudooverbo/ Pseudoadjetivo		
<b>A</b>	Nossa/Essa	x-o vato	y-[iw] muniu	y-[iw] munil	
	Meu/Seu	x-e dale	y-[iw] biliu	y-[iw] bilil	
<b>B</b>	Nossa/Essa	x-e lepe	y-[iw] pigiu	y-[iw] pigil	
	Meu/Seu	x-o pifo	y-[iw] cutiu	y-[iw] cutil	

Tabela 5 – Visualização das combinações entre Det e terminação de N nas fases de familiarização.

#### 4- Pós teste:

Após a apresentação dos 12 ensaios da fase de teste, um estímulo auditivo, com duração de aproximadamente 7 segundos, foi apresentado sequencialmente para a criança. O tempo de escuta desse áudio também foi computado e posteriormente comparado com o tempo de escuta dos *ensaios* testes, a fim de garantir que a criança tenha se mantido atenta até o final da atividade. Entretanto, não foram utilizados para o cálculo final da média das condições. Assim como na fase de habituação, uma pequena história com prosódia dirigida à criança era apresentada:

“Oba! Que bom que você veio! Adorei brincar com você. Você quer ser meu amigo?”

Imagen das fases habituação, familiarização, teste e pós teste:



Figura 5 – Imagem Familiarização e Teste

Imagen “attention getter”:

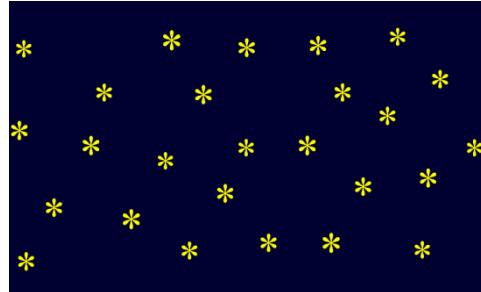


Figura 6 – Imagem “attention getter”

### O procedimento

O responsável, ao chegar com o bebê, em horário pré-agendado, é recebido por dois pesquisadores na antessala do laboratório. Trata-se de um ambiente aconchegante tanto para os responsáveis quanto para os bebês. Há, neste ambiente, poltronas, uma mesa, um tapete emborrachado colorido e alguns brinquedos. Neste momento algumas orientações gerais sobre o bom aproveitamento da atividade são fornecidas, como por exemplo: não falar com a criança durante a atividade, não virar o corpinho ou a cabeça da criança em direção ao monitor, entre outros, mas não é revelado o objetivo da atividade.

Assim que verificamos que a criança já se ambientou ao recinto e, nesta hora, estando presente apenas um pesquisador, enquanto o outro já se encontra posicionado na sala de controle, o responsável é convidado a entrar na sala de experimento onde se encontram “visíveis” apenas uma poltrona e um monitor (distância de 1,5 m entre poltrona e monitor). Além desses dois equipamentos, encontram-se, abaixo do monitor, uma câmera que fornece as imagens ao experimentador e uma caixa de som amplificada, coberta embaixo da toalha da mesa.

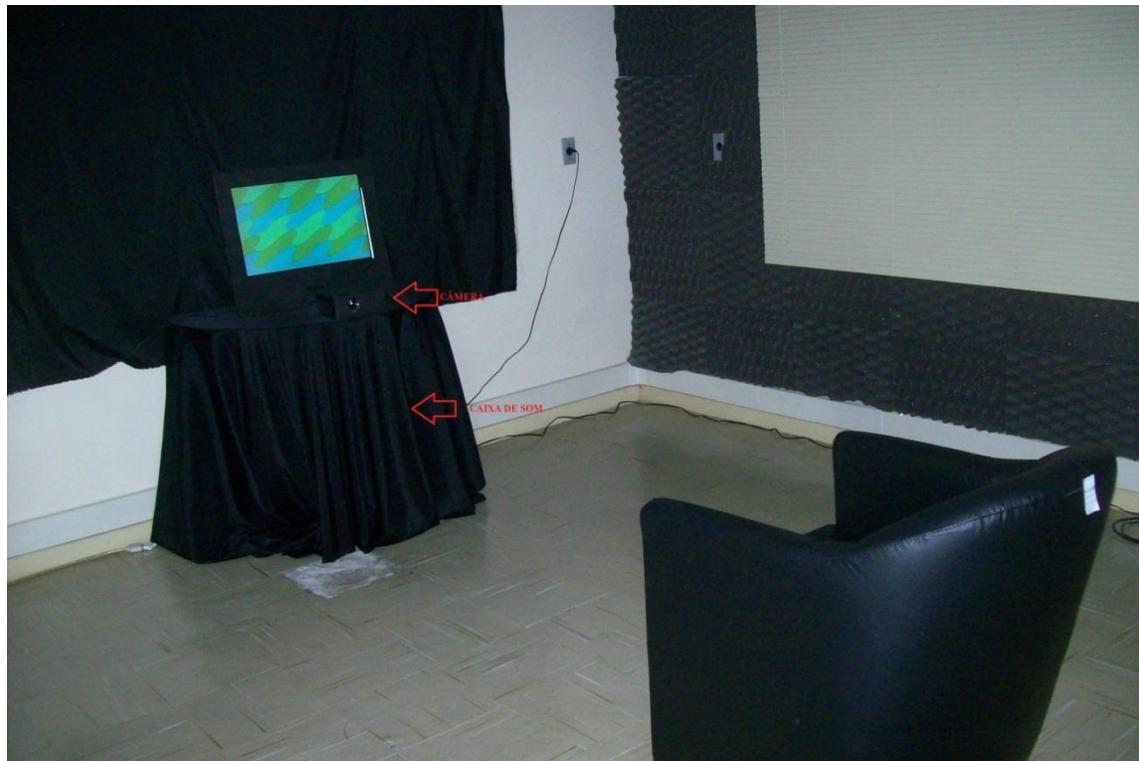


Figura 7 - Cabine de atividades experimentais onde as atividades foram realizadas na UFJF.

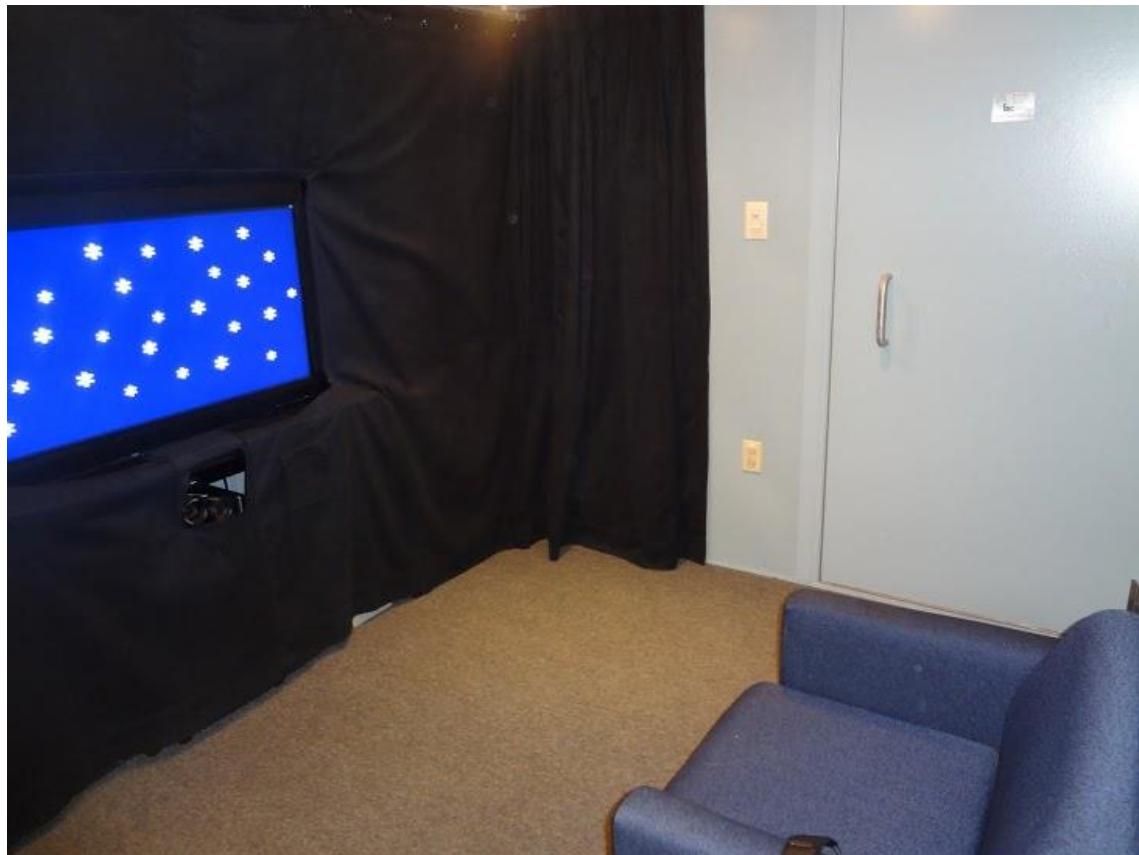


Figura 8 - Cabine de atividades experimentais onde as atividades foram realizadas na UQAM.

Assim que responsável e bebê se acomodam no ambiente, um dos experimentadores coloca um fone de ouvido no responsável, com uma música clássica, a fim de evitar que ouça os estímulos auditivos da atividade e interfira na atividade, mesmo de forma inconsciente. O experimentador assistente se retira da sala e o experimentador que marcará o tempo de olhar da criança inicia a execução do programa. Vale ressaltar que o experimentador foi o mesmo em todas as atividades realizadas com todas as crianças, não existindo, portanto, diferenças em relação ao modo de aplicação da técnica. O experimentador, através da visualização do olhar da criança, que é disponibilizado a partir de um monitor instalado na cabine de observação, “registra”, através do software *Habit* e do computador, o “tempo de olhar” de cada estímulo.

Após a familiarização, iniciam-se os dois minutos de habituação. Como dito anteriormente, nesta fase o tempo de fixação do olhar da criança não é computado, visto que interessa-nos que a criança esteja ouvindo “naturalmente” a pseudolíngua que é apresentada. Após os dois minutos de familiarização, a tela do “attention getter” aparece, permitindo ao pesquisador saber que a fase teste irá começar (pois há um contraste de luminosidade das imagens apresentadas entre as fases de habituação e teste e o “attention getter”). O pesquisador inicia a fase teste e faz a marcação do tempo de fixação do olhar da criança ao estímulo. Se a criança fixar o olhar para o centro da tela, o tempo é marcado e caso desvie o olhar, o tempo deixa de ser marcado. O mesmo ocorre na fase postteste. Após o postteste, a atividade é finalizada e responsável e bebê são conduzidos até a antessala. A duração média do experimento é de 5 minutos.

#### **4.2.5 -Resultados e Discussões (1A)**

Os resultados obtidos mostram que bebês canadenses ouviram por mais tempo a condição agramatical tanto na condição Adjetivo quanto na condição Verbo, o que evidencia capacidade de discriminar os estímulos apresentados. A média do tempo de escuta na condição Adjetivo foi 7,21 segundos na condição grammatical e 7,28 segundos na condição agramatical, enquanto que a média do tempo de escuta na condição Verbo foi 5,99 segundos na condição grammatical e 7,10 segundos na condição agramatical.

O tempo médio de escuta para cada condição pode ser visualizado no gráfico abaixo<sup>15</sup>:

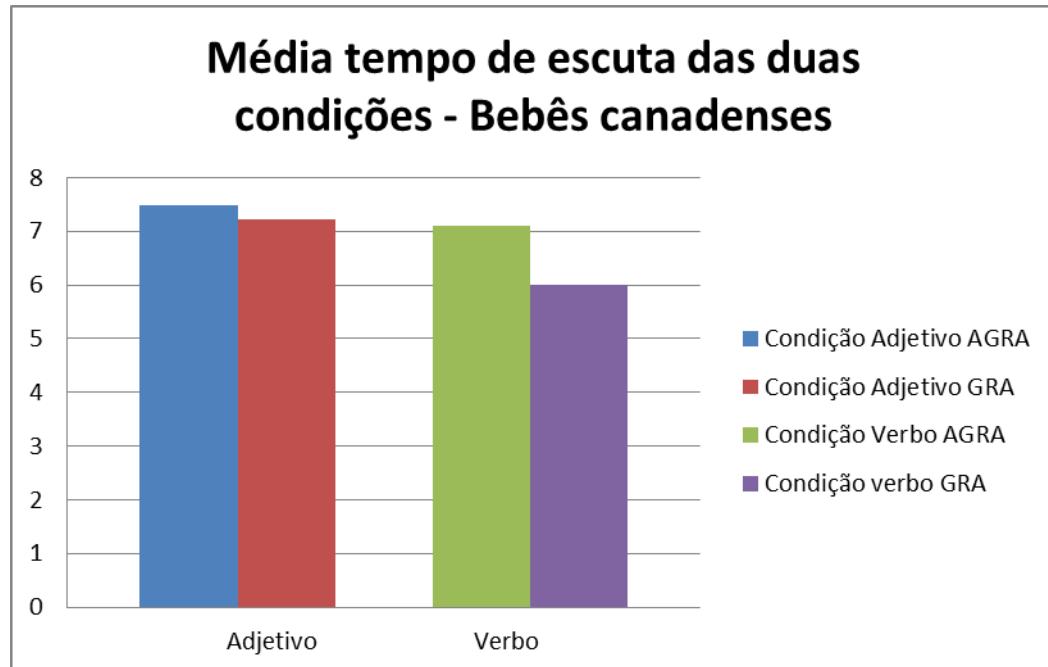


Gráfico 2 – Resultado com bebês canadenses- Tempo médio em cada condição

Antes de realizarmos a ANOVA, um teste de normalidade<sup>16</sup> dos dados das duas condições foi realizado, evidenciando que nossas amostras estavam normais, como podemos visualizar nas tabela abaixo. A primeira relata a significância do teste de normalidade em cada condição e a segunda, a aleatoriedade dos pontos (amostra) em torno da reta nas duas condições:

---

<sup>15</sup> Para o tempo individual de cada criança, nas duas condições gramaticais, ver Anexo 2.

<sup>16</sup> Os **testes de normalidade** são usados para determinar se um conjunto de dados de uma dada variável aleatória é bem modelada por uma distribuição normal ou não, ou para calcular a probabilidade da variável aleatória subjacente estar normalmente distribuída. O teste utilizado em nossas amostras foi o Shapiro-Wilk e o programa utilizado foi o SPSS 2.0. Se o valor do teste de normalidade for superior a **0,05**, as variáveis em estudo seguem uma distribuição normal. Esse fato é confirmado pela aleatoriedade dos pontos em torno da reta.

Teste de normalidade – Shapiro Wilk		Statistic	Df	Sig.
Condição Adjetivo	GRAMATICAL	,942	16	,372
	AGRAMATICAL	,926	16	,207
Condição Verbo	GRAMATICAL	,917	16	,152
	AGRAMATICAL	,956	16	,599

Tabela 6 – Teste de normalidade das amostras das condições Adjetivo e Verbo – Experimento 1A

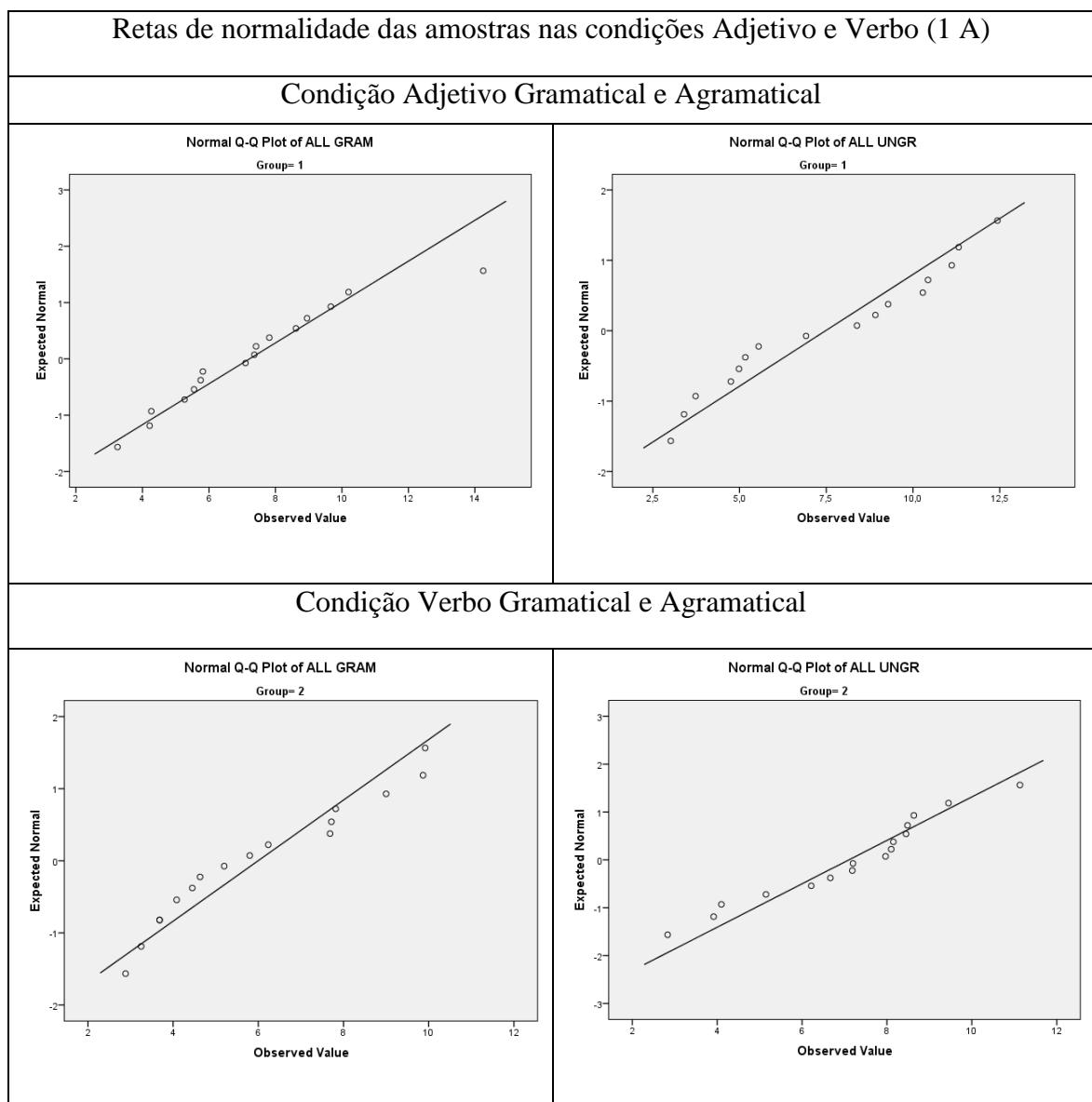


Tabela 7 - Retas de normalidade das condições Adjetivo e Verbo- Experimento 1A

Os resultados obtidos foram submetidos a uma análise da variância (ANOVA) com design fatorial 2X2, tendo a variável *Alinhamento das DNA às fronteiras prosódicas* (condição Verbo vs. condição Adjetivo) como fator *between-subjects*, e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal ou interação.

Os resultados da condição Verbo foram submetidos a uma ANOVA com design 2X2, apresentando a variável *combinação de padrão da fase de familiarização* como fator *between-subjects* e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal de *combinação de padrão da fase de familiarização* e efeito de interação. Entretanto, houve efeito significativo principal de gramaticalidade  $F(1,14) = 5,09$   $p<0,0405$ , evidenciando que bebês diferenciaram os ensaios gramaticais e agramaticais, escutando por um tempo maior os ensaios agramaticais, independente do padrão ao qual foram familiarizados.

Os resultados da condição Adjetivo também foram submetidos a uma ANOVA com design 2X2, apresentando a variável *combinação de padrão da fase de familiarização* como fator *between-subjects* e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal ou interação.

Análises estatísticas adicionais foram realizadas, buscando analisar os resultados em conjunto de cada condição, independente do padrão com o qual o bebê foi familiarizado. Um teste estatístico *T-Student* foi realizado nas duas condições e o resultados mostraram um *p* não significativo para a condição Adjetivo ( $t(15) 0,95$ ,  $p= 0,35$ ) e *p* significativo para a condição Verbo ( $t(15) 2,62$ ,  $p= 0,018$ ). O tamanho do efeito também foi calculado, assim o tamanho do efeito para a condição Adjetivo é  $d= 0,08$  enquanto que para a condição Verbo é  $d= 0,32$ <sup>17</sup>. As

---

<sup>17</sup> O tamanho de efeito é uma estatística descritiva que serve como complemento ao teste de significância estatística. Essa medida descreve os efeitos observados: efeitos grandes, mas não significantes, podem sugerir que as pesquisas futuras necessitam de maior poder, enquanto efeitos pequenos, mas significantes devido ao grande tamanho amostral, podem levar a uma supervalorização do efeito observado. Cohen (1988) apud Lindenau e Guimarães (2012) sugeriu alguns pontos de corte para classificação do tamanho de efeito. Valores superiores ou iguais a 0,8 representam tamanho de efeito grande; entre 0,8 a 0,2 são considerados médios e inferiores a 0,2 pequenos. Contudo, esses valores podem variar conforme a área de estudo, devendo ser utilizados somente quando há uma base para estimar a classificação do tamanho de efeito para o conjunto de dados que se está trabalhando. Sendo assim, se seguirmos a linha proposta por Cohen devemos comparar nossos efeitos com aqueles previamente estabelecidos dentro de nossa área de investigação, pois obter um efeito de 0,5 quando os valores tipicamente observados são de 0,2, pode significar um efeito importante. Como na área de aquisição da linguagem, não é usual relatar esse tipo de análise, não podemos estabelecer comparações em relação ao tamanho do efeito. Para mais informações, consultar Lindenau e Guimarães (2012).

análises evidenciam que, apesar de haver diferença entre o tempo de escuta entre os ensaios gramaticais e agramaticais nas duas condições, obtivemos uma diferença estatisticamente significativa somente entre os ensaios gramaticais e agramaticais na condição Verbo.

Como vimos, a única diferença entre as duas condições era a fronteira de sintagma fonológico alinhada à terminação do Nome (à DNA). Dessa forma, os resultados indicam que essa diferença foi suficiente para que os bebês canadenses conseguissem identificar, abstrair e generalizar as DNAs a novos estímulos, diferentemente do que ocorreu na condição Adjetivo.

Assim, nota-se que a presença de fronteira de sintagma fonológico é um característica/propriedade perceptualmente disponível para os bebês e que, quando alinhada à DNA, auxilia na identificação, abstração e generalização desses padrões. Além disso, esses resultados nos fornecem evidências da importância em se pensar em uma integração dessas pistas, pois somente DNAs (condição Adjetivo), nesse contexto experimental, não forneceu pistas suficientes em comparação às DNAs integradas à presença de fronteira de sintagma fonológico (condição Verbo) para a abstração e generalização dos padrões de DNA. Destaca-se novamente que os bebês canadenses estavam diante de estímulos completamente novos, diferentes de sua língua, evidenciando sua rápida capacidade de segmentação do *continuum* sonoro, reconhecimento dos itens funcionais e mapeamento dos itens do léxico.

Diante desses resultados, uma indagação nos levou a expandir essa atividade: até que ponto esses resultados decorrem simplesmente da novidade, da apresentação de estímulos do “tipo” língua, mas de uma língua diferente daquela a que os bebês, ao final do primeiro ano de vida, já estão familiarizados? Decidimos, dessa forma, testar bebês brasileiros, visto que eles já estão familiarizados com os itens de sua língua, mas estes agora – durante a atividade – aparecem em uma nova configuração. Seria a pista de fronteira de sintagma fonológico suficiente para o bebê identificar e abstrair novos padrões, semelhantes porém distintos daqueles de sua língua?

#### **4.2.6- Experimento prosódia bebês brasileiros (1B)**

As variáveis, condições experimentais e estímulos foram mantidos.

**Hipótese:**

Elementos relacionados entre si em uma configuração de DNAs alinhados a fronteiras de sintagma fonológico são mais facilmente percebidos, abstraídos e generalizados por bebês brasileiros aos 11 meses de idade, mesmo quando esses elementos são apresentados em uma nova configuração em relação a sua língua de origem, do que quando ao menos um desses elementos não se encontra em borda de fronteira de sintagma fonológico.

## **Previsão**

Se DNAs alinhadas a bordas de fronteiras prosódicas são mais facilmente percebidas, abstraídas e generalizadas do que quando ao menos um desses elementos não se encontra em borda de fronteira prosódica, bebês brasileiros com média de idade de 11 meses serão capazes ou terão uma melhor performance ao abstrair o padrão e generalizá-lo em novos estímulos na condição alinhamento dos elementos de DNAs a bordas de sintagmas fonológicos (Condição Verbo), do que na condição em que os elementos de DNAs não estão alinhados a bordas de fronteiras de sintagma fonológico (Condição Adjetivo), mesmo quando esses elementos são apresentados em uma nova configuração em relação à sua língua de origem.

## **Participantes brasileiros**

Participaram da atividade experimental 43 bebês brasileiros<sup>18</sup> com idade média de 11 meses<sup>19</sup> (10 meses e 0 dias a 12 meses e 29 dias). Entretanto, 11 bebês foram eliminados por inquietação (4), choro (6) e problemas técnicos durante a sessão (1). Assim, os resultados encontrados são referentes a 32 bebês, dentre os quais 13 são do sexo feminino e 19 são do sexo masculino. Para assegurar a confiabilidade dos resultados, os 32 sujeitos foram divididos em dois grupos. Dessa forma, 16 crianças constituem o Grupo A- Condição Adjetivo - (6 do sexo feminino e 10 do sexo masculino) e 16 constituem o Grupo B – Condição Verbo- (7 do sexo feminino e 9 do sexo masculino). Os participantes foram recrutados pelos pesquisadores do NEALP/UFJF e não receberam nenhuma retribuição financeira e/ou de outra ordem pela participação na atividade. Os responsáveis foram informados que a atividade é aprovada pelo

---

<sup>18</sup> Bebês brasileiros da cidade de Juiz de Fora-MG.

<sup>19</sup> Devido a uma dificuldade maior no recrutamento de crianças, optamos pela ampliação da faixa etária, mantendo, no entanto, a idade média de 11 meses.

Comitê de Ética da UFJF e foram orientados a preencher um formulário - Cadastro de Crianças<sup>20</sup> - com as informações sobre o bebê e assinar um Termo de Consentimento livre e esclarecido<sup>21</sup>, caso estivessem de acordo em participar da pesquisa.

#### **4.2.6.1- Resultados e discussões (1B)**

Os resultados obtidos mostram que bebês brasileiros, assim como os canadenses, ouviram por mais tempo a condição agramatical tanto na condição Adjetivo quanto na condição Verbo, o que evidencia capacidade de discriminar os estímulos apresentados. A média do tempo de escuta na condição Adjetivo foi 6,39 segundos na condição gramatical e 7,11 segundos na condição agramatical, enquanto que a média do tempo de escuta na condição Verbo foi 6,40 segundos na condição gramatical e 8,36 segundos na condição agramatical.

Antes de realizarmos a ANOVA, assim como para o grupo anterior, um teste de normalidade dos dados das duas condições foi realizado, evidenciando que nossas amostras estavam normais, como podemos visualizar nas tabelas abaixo. A primeira relata a significância do teste de normalidade em cada condição e a segunda, a aleatoriedade dos pontos (amostra) em torno da reta nas duas condições:

Teste de normalidade – Shapiro Wilk		Statistic	Df	Sig.
Condição Adjetivo	GRAMATICAL	,922	16	,180
	AGRAMATICAL	,912	16	,125
Condição Verbo	GRAMATICAL	,964	16	,730
	AGRAMATICAL	,953	16	,540

Tabela 8 -Teste de normalidade das amostras das condições Adjetivo e Verbo – Experimento 1B

---

<sup>20</sup>Anexo 3

<sup>21</sup>Anexo 4

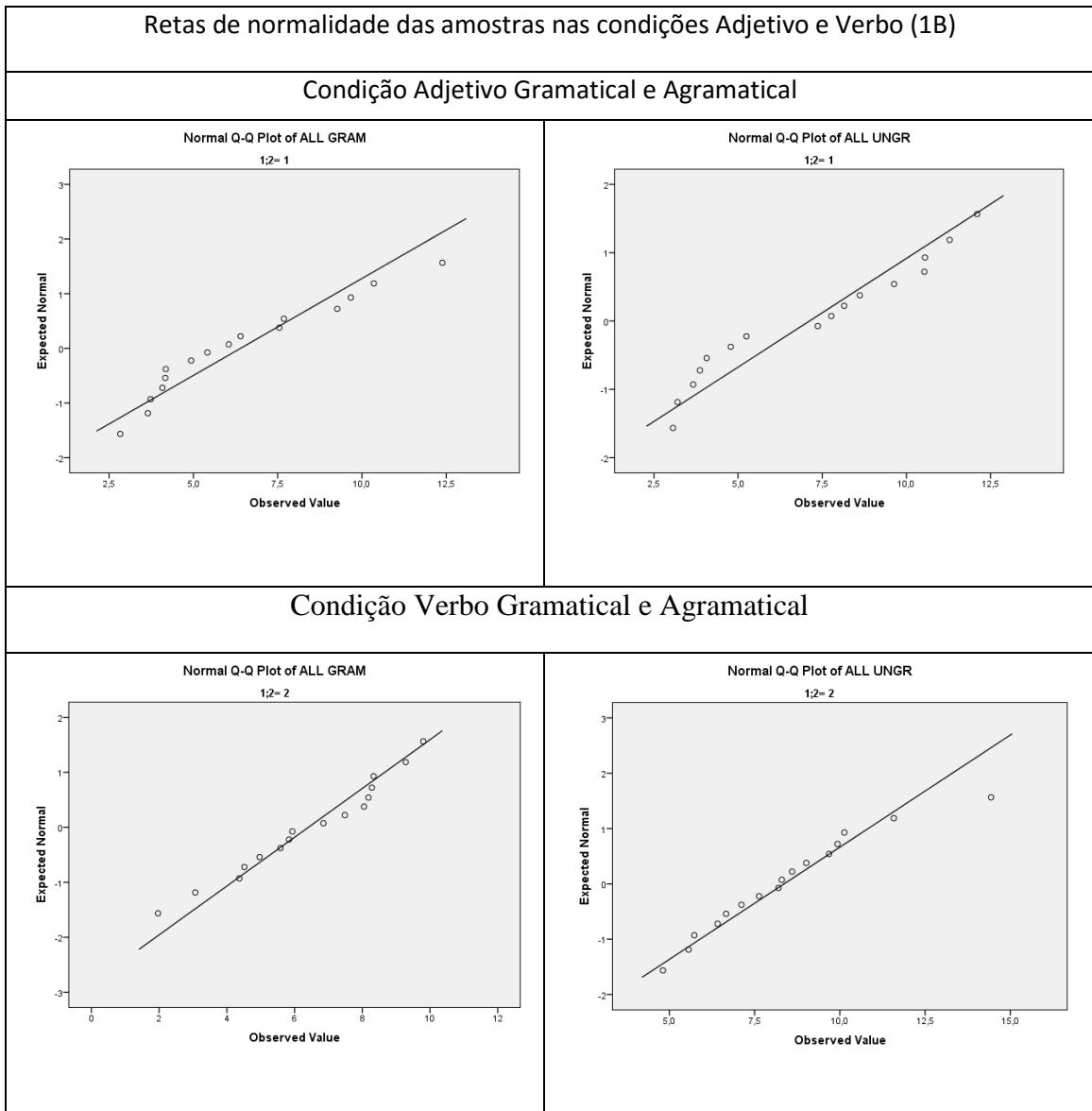


Tabela 9 - Retas de normalidade das condições Adjetivo e Verbo- Experimento 1B

O tempo médio de escuta para cada condição pode ser visualizado no gráfico abaixo<sup>22</sup>:

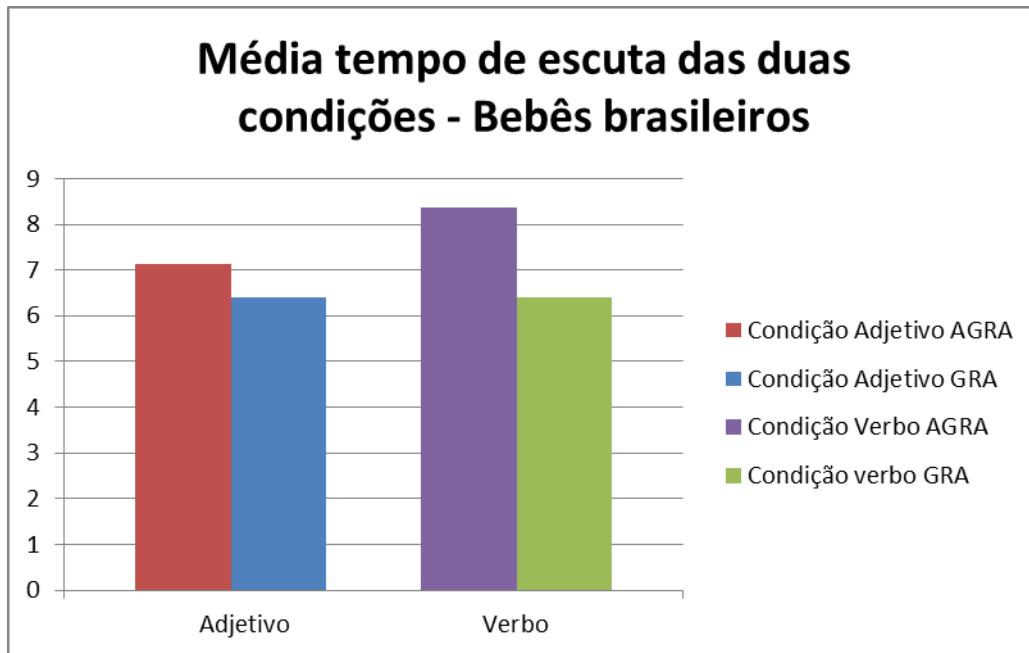


Gráfico 3 – Resultado com bebês brasileiros- Tempo médio em cada condição

Os resultados obtidos foram submetidos a uma análise da variância (ANOVA) com design fatorial 2X2, tendo a variável *Alinhamento das DNA às fronteiras prosódicas* (condição Verbo vs. condição Adjetivo) como fator *between-subjects*, e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal em relação ao alinhamento das DNAs às fronteiras prosódicas ou interação. Entretanto, houve efeito de gramaticalidade em relação à familiarização  $F(1,30) = 10,8$ ,  $p<0,002609$ .

Os resultados da condição Verbo foram submetidos a uma ANOVA com design 2X2, apresentando a variável *combinação de padrão da fase de familiarização* como fator *between-subjects* e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal de *combinação de padrão da fase de familiarização* e efeito de interação. Entretanto, houve efeito significativo principal de gramaticalidade  $F(1,14) = 14,7$ ,  $p<0,0018$ , evidenciando que bebês diferenciaram os ensaios gramaticais e agramaticais, escutando por um tempo maior os ensaios agramaticais, independente do padrão ao qual foram familiarizados.

<sup>22</sup> Para o tempo individual de cada criança, nas duas condições gramaticais, ver Anexo 5.

Os resultados da condição Adjetivo também foram submetidos a uma ANOVA com design 2X2, apresentando a variável *combinação de padrão da fase de familiarização* como fator *between-subjects* e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. Não foram encontrados efeito principal ou interação.

Análises estatísticas adicionais foram realizadas, buscando analisar os resultados em conjunto de cada condição, independente do padrão com o qual o bebê foi familiarizado. Um teste estatístico *T-Student* foi realizado nas duas condições e o resultados mostraram um p não significativo para a condição Adjetivo ( $t(15) = 1,61$ ,  $p = .1268$ ) e p significativo para a condição Verbo ( $t(15) = 3,93$ ,  $p = .0013$ ). As análises evidenciam que, apesar de haver diferença entre o tempo de escuta entre os ensaios gramaticais e agramaticais nas duas condições, obtivemos uma diferença estatisticamente significativa somente entre os ensaios gramaticais e agramaticais na condição Verbo.

Ainda que os resultados com crianças brasileiras, assim como com as canadenses, indiquem um tempo maior de escuta nos ensaios agramaticais em ambas as condições (Adjetivo 6,39 segundos na condição grammatical e 7,11 segundos na condição agramatical e Verbo 6,40 segundos na condição grammatical e 8,36 segundos na condição agramatical), a diferença entre as médias dos estímulos gramaticais e agramaticais só foram estatisticamente significativas na condição Verbo. Considerando que a única diferença entre as duas condições era a fronteira de sintagma fonológico alinhada à terminação do Nome (à DNA) na condição Verbo, os resultados indicam que essa diferença foi suficiente para que os bebês brasileiros conseguissem identificar, abstrair e generalizar as DNAs a novos estímulos, diferentemente do que ocorreu na condição Adjetivo.

Sabemos que esse experimento é diferente para os bebês brasileiros, em relação aos canadenses, por já conhecerem os determinantes, o contorno prosódico característico e a terminação dos pseudoadjetivos/pseudoverbos. Entretanto, não podemos considerar tal fato uma vantagem; ao contrário, essa tarefa pode ser um mais difícil para eles. Para os bebês canadenses, toda a atividade era construída de elementos novos/não conhecidos, dessa forma, a tarefa deles era reconhecer, abstrair e generalizar um novo padrão nunca antes ouvido. Para os bebês brasileiros, entretanto, a tarefa consiste em reconhecer e abstrair um novo padrão (em uma nova subclassificação) distinto dos padrões já (re)conhecidos de sua língua materna. Ao invés de “aprender” um novo padrão, os bebês brasileiros tiveram que subcategorizar novamente uma padrão possivelmente já aprendido anteriormente.

#### **4.3- Atividade experimental 2 – Capacidade de categorização**

Conforme dito, essa atividade surgiu a partir de resultados anteriores de Teixeira & Name (2012) que sugerem que bebês brasileiros, aos 11 meses de idade, a partir de um curto período de familiarização (2 minutos) são capazes de identificar, abstrair e generalizar padrões de DNA entre o Det e terminação de N. A partir desses resultados, surgiu a indagação se as crianças seriam capazes de categorizar elementos lexicais a partir dos padrões de DNA.

Como vimos no capítulo 2, DNAs são fundamentais para o processo de aquisição de uma língua natural e um dos fatores para sua importância deve-se ao fato de DNAs expressarem a (sub)categorização de novas palavras, ainda que o material interveniente seja bastante variável, como a subclassificação de N masculinos e femininos no português. Em nossa atividade desenvolvida anteriormente (TEIXEIRA & NAME, 2012), não foi possível defender que os bebês foram capazes de categorizar os itens do léxico, assim como não há na literatura essa investigação.

Nosso objetivo, portanto, ao propormos essa atividade não é só investigar a capacidade de abstração e generalização de padrões linguísticos por bebês durante o primeiro ano de vida, mas, principalmente, investigar os fatores facilitadores desse processo, assim como o uso de tais capacidades na categorização inicial de elementos do léxico. A partir do ambiente linguístico da atividade desenvolvida anteriormente, ou seja, o DP, buscamos condições em que pudéssemos testar a categorização dos elementos do léxico a partir do reconhecimento, abstração e generalização de DNAs. Uma das opções encontradas e que ocorre no PB foi testar a forma do Det com a terminação do N, caracterizando uma relação de DNA, e avaliar em que medida bebês agrupam determinantes em uma mesma categoria a partir da identificação, abstração e generalização dessa relação de DNA.

Na atividade experimental proposta, utilizamos a mesma técnica da atividade experimental anterior – Olhar preferencial, e as atividades foram desenvolvidas no laboratório do NEALP/UFJF. Como essa técnica já foi apresentada anteriormente, passaremos, a seguir, para a elaboração dos estímulos.

#### **4.3.1- Elaboração dos estímulos**

O ambiente escolhido para essa atividade foi o DP. Dessa forma, utilizamos quatro Det masculinos do português: “Seu, Meu, Nosso, Esse” e pseudonomes dissílabos e paroxítonos.

Para a fase de familiarização – com duração de 2 minutos contínuos –, utilizamos 8 pseudonomes paroxítonos terminados em *-e*, 8 pseudonomes paroxítonos terminados em *-a*<sup>23</sup>, 4 pseudonomes oxítonos terminados em *-u* e 4 pseudonomes oxítonos terminados em *-é* [ɛ]. Para um maior controle, dividimos os participantes em dois grupos: No Grupo A, os pseudonomes terminados em *-u* eram combinados com os determinantes “Nosso” e “Seu” e a terminação *-e* combinada com o determinante “Seu”; a terminação *-é* eram precedida pelos determinantes “Esse” e “Meu” e os pseudonomes terminados em *-a* precedidos somente pelo determinante “Meu”. No Grupo B, o padrão era o oposto. As combinações dos Det e terminação dos N dos dois grupos, podem ser melhor visualizadas abaixo, assim como os estímulos utilizados podem ser visualizados na tabela a seguir:

---

<sup>23</sup>Os 16 pseudonomes (8 pseudonomes terminados em *-e*, 8 pseudonomes terminados em *-a*) foram os mesmos utilizados na atividade de TEIXEIRA & NAME (2012).

Grupo A: seu +**X-e**; seu/nosso + **X-u**; meu +**X-a**; meu/esse + **X-E**  
 Grupo B: seu+**X-a**; seu/nosso + **X-E**; meu+**X-e**; meu/esse + **X- u**

<b>DPs: Det + X-e or + X-u Nome</b>				<b>DPs: Det + X-a or X-E Nome</b>			
Determinante		Pseudopalavra		Determinante		Pseudopalavra	
<b>Grupo A</b>	<b>Grupo B</b>	Bafe ['bafí] Mepe ['mepi] Guque ['guki] Pôfe ['povi] Dabe ['dabi] Quive ['kivi] Zobe ['zòbi] Nêque ['neki]		<b>Grupo A</b>	<b>Grupo B</b>	Ziba ['ziba] Tafa ['tafa] Nôca ['noka] Vêba ['veba] Bapa ['bapa] Goca ['gôka] Mefa ['mefa] Puva ['puva]	
Seu ['sew]	Meu ['mew]	--		Meu ['mew]	Seu ['sew]	-	
-	--	Bechu [be'su] Tadu [ta'du] Jomu [žo'mu] Chivu [ši'vu]		Esse ['esi]	Nosso ['nòsu]	Sogué [so'ge] Pilé [pi'lε] Jufé [žu'fe] Maké [ma'ke]	
Seu ['sew]	Meu ['mew]						
Nosso ['nòsu]	Esse ['esi]						

Tabela 10 - Esquema das combinações utilizadas na atividade

Para a fase de teste, foram utilizados dois determinantes da fase de familiarização, “esse” e “nosso”, e quatro novos pseudonomes paroxítonos que não haviam aparecido na fase de familiarização: “mufe, tova, guika e pebe” (2 com terminação –e e 2 com terminação –a). A partir da combinação desses Det e pseudonomes, foram criados DPs para os ensaios da fase teste e cada ensaio apresenta 6 repetições do mesmo DP, com duração total de 16 s em cada ensaio. Quatro ensaios diferentes compunham a fase teste, totalizando 12 ensaios na fase, visto que cada ensaio aparecia três vezes na fase teste da atividade.

Dessa forma, no Grupo A, por exemplo, apesar de não ter ouvido a combinação “**Nosso pebe**” durante a fase de familiarização, se a criança for capaz de categorizar os elementos a partir do reconhecimento, abstração e generalização das DNAs, ela irá agrupar o “nosso” com o “seu”, já que o “seu” foi combinado com a terminação –e na fase de familiarização. A tabela com todos os estímulos está disponível no anexo 6.

### **4.3.2- Variáveis experimentais**

**Variável independente:** Padrão de DNAs (combinação entre Determinantes e pseudonomes).

2 níveis: 1) congruente (em relação aos padrões familiarizados previamente).

2) incongruente (em relação aos padrões familiarizados previamente).

**Variável dependente:** tempo de escuta/olhar.

#### **Condições Experimentais:**

- GRAMATICAL: Os ensaios da fase de teste são congruentes com o padrão apresentado na fase de familiarização.
- AGRAMATICAL: Os ensaios da fase de teste são incongruentes com o padrão apresentado na fase de familiarização.

#### **Hipótese:**

Bebês, com média de idade de 11 meses, são capazes categorizar os elementos do léxico a partir da identificação, abstração e generalização do padrão de DNA entre Det e terminação do N.

**Previsão:** Se bebês, com média de idade de 11 meses, são capazes de categorizar os elementos do léxico a partir da identificação, abstração e generalização do padrão de DNA entre Det e terminação do N, mesmo não tendo escutado uma determinada combinação entre Det e terminação de N, agrupará os Dets em uma mesma categoria (visto que eles foram familiarizados com uma outra terminação na fase de familiarização).

## **Participantes**

Participaram da atividade experimental 25 bebês brasileiros<sup>24</sup> com idade média de 11 meses<sup>25</sup> (10 meses e 0 dias a 12 meses e 29 dias). Entretanto, 7 bebês foram eliminados por inquietação (3) e choro (4). Assim, os resultados encontrados são referentes a 18 bebês, dentre os quais 7 são do sexo feminino e 11 são do sexo masculino. Para assegurar a confiabilidade dos resultados, os 18 sujeitos foram divididos em dois grupos. Dessa forma, 9 crianças constituem o Grupo A (5 do sexo feminino e 4 do sexo masculino) e 9 constituem o Grupo B (2 do sexo feminino e 7 do sexo masculino). Os participantes foram recrutados pelos pesquisadores do NEALP/UFJF e não receberam nenhuma retribuição financeira e/ou de outra ordem pela participação na atividade. Os responsáveis foram informados que a atividade é aprovada pelo Comitê de Ética da UFJF, e foram orientados a preencher um formulário - Cadastro de Crianças - com as informações sobre o bebê e a assinar um Termo de Consentimento livre e esclarecido, caso concordassem com a atividade.

O Material utilizado, as fases da atividade experimental e os equipamentos foram os mesmos da atividade anterior.

### **4.3.3- Resultados e Discussões**

A média do tempo de escuta nos mostra que os bebês ouviram por mais tempo os ensaios agramaticais (7,95 segundos) em relação aos ensaios gramaticais (6,70 segundos)<sup>26</sup>, sugerindo que bebês foram capazes de discriminar os ensaios do teste.

Antes de realizarmos a ANOVA, um teste de normalidade dos dados das duas condições foi realizado, evidenciando que nossas amostras estavam normais, como podemos visualizar nas tabela abaixo. A primeira relata a significância do teste de normalidade e a segunda, a aleatoriedade dos pontos (amostra) em torno da reta nas duas condições:

---

<sup>24</sup> Bebês brasileiros da cidade de Juiz de Fora.

<sup>25</sup> Devido a uma dificuldade maior no recrutamento de crianças, optamos pela ampliação da faixa etária, mantendo, no entanto, a idade média de 11 meses.

<sup>26</sup> Para o tempo individual de cada criança, nas duas condições gramaticais, ver Anexo 7.

Teste de normalidade – Shapiro Wilk	Statistic	Df	Sig.
GRAMATICAL	,918	20	<b>,089</b>
AGRAMMATICAL	,932	20	<b>,165</b>

Tabela 11– Teste de normalidade das amostras do Experimento de categorização

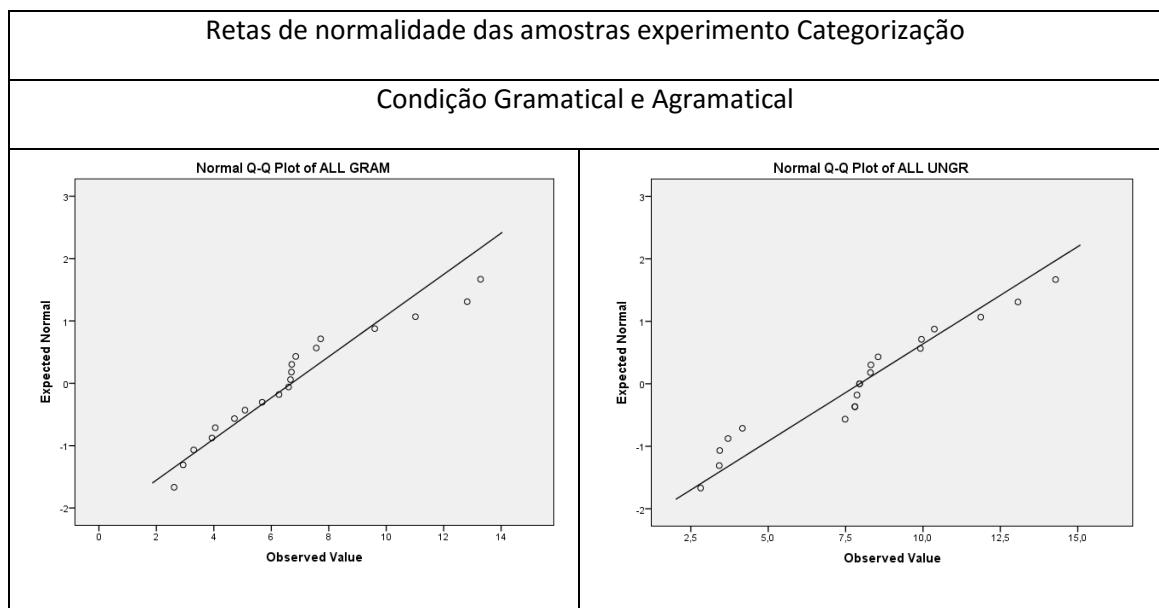


Tabela 12- Retas de normalidade experimento Categorização

O tempo médio de escuta para cada condição (gramatical e agramatical) pode ser visualizado no gráfico abaixo:

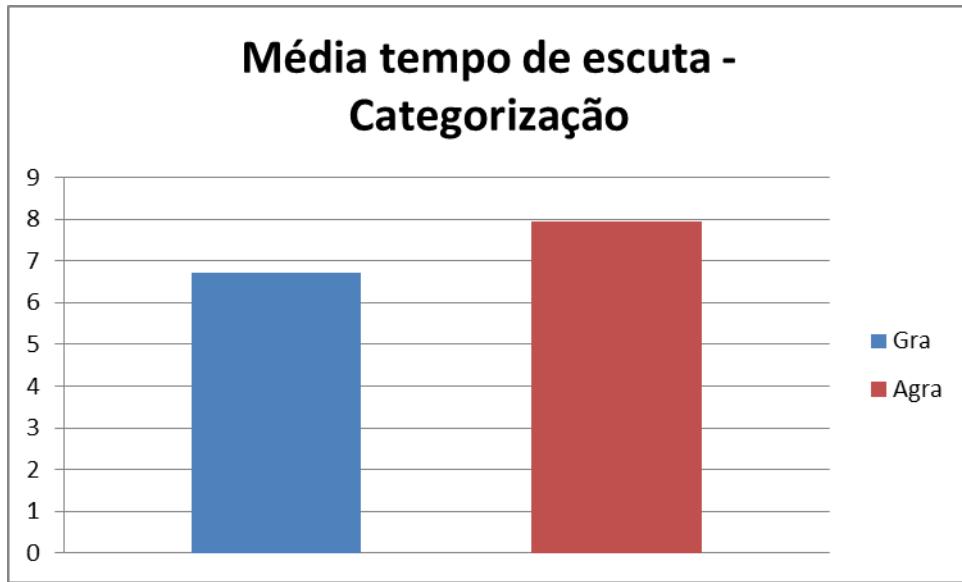


Gráfico 4 – Resultado média tempo de escuta experimento categorização

Os resultados obtidos foram submetidos a uma análise da variância (ANOVA) com design factorial 2X2, tendo a variável *Padrão de familiarização* como fator *between-subjects*, e *Gramaticalidade em relação à familiarização* (condição Gramatical vs. condição Agramatical) como fator *within-subjects*. O único efeito encontrado foi o de gramaticalidade em relação à familiarização  $F(1,14) = 5,07$ ,  $p < 0,0409$ .

Um teste estatístico *T-Student* foi realizado, independente do padrão com o qual o bebê foi familiarizado, e os resultados mostraram um *p* significativo entre as condições grammatical e agramatical  $t(17)=3,12$ ,  $p= 0,006$ . O tamanho do efeito também foi calculado  $d= 0,253$ .

As análises evidenciam que os dois minutos de familiarização parecem ter sido suficientes para permitir a categorização dos elementos a partir da combinação entre Determinantes e terminação de Nomes. Dessa forma, supomos que a abstração e generalização do padrão de DNAs também parecem estar presente, visto que seria “requisito”, nesta tarefa especificamente, para as habilidades iniciais de categorização.

Essa atividade experimental também foi aplicada com bebês canadenses em período concomitante ao dos bebês brasileiros, como parte de pesquisas desenvolvidas pela orientadora desta tese durante seu estágio de pós doutorado. Os resultados indicam que bebês canadenses (média 11 meses de idade) foram capazes de discriminar os ensaios ouvidos,

escutando por mais tempo os ensaios gramaticais 6,06 s em relação aos ensaios agramaticais 5,27 s. A diferença entre o tempo médio de escuta dos ensaios se mostrou significativa:  $F(1,30)=6,69$   $p<.015$ . Os resultados sugerem que bebês canadenses aos 11 meses, mesmo expostos a estímulos totalmente diferentes de sua língua materna (i) foram capazes de abstrair regras de terminação de N que não estavam explícitas no estímulo e (ii) utilizando os padrões de DNAs, determinaram categorias gramaticais de novas palavras.

Assim como na atividade experimental anterior, aplicar a mesma atividade em dois grupos de participantes distintos nos permite fazer ao menos duas conclusões: i) bebês são muito eficientes no tratamento do estímulo linguístico, sendo capazes de reconhecer, abstrair e generalizar padrões de DNAs mesmo a partir de um curto período de exposição (bebês brasileiros e canadenses) e em uma língua totalmente diferente de sua língua materna (bebês canadenses) e igualmente capazes de categorizar itens do léxico a partir das relações de DNAs; ii) a tarefa dos bebês brasileiros parece ser mais difícil que para os bebês canadenses, visto que enquanto para estes todos os elementos são novos, aqueles já conhecem os determinantes e a subclassificação e tiveram que em um tempo curto abstrair uma nova subclassificação e abstraí-la a novos ensaios.

Ao avaliarmos o conjunto de resultados dos nossos experimentos, verificamos que na atividade experimental 1, tanto bebês canadenses quanto bebês brasileiros apresentaram média de escuta maior para ensaios agramaticais em ambas as condições (Condição Adjetivo e Condição Verbo). Na atividade experimental 2, bebês brasileiros apresentaram uma média de escuta maior para ensaios agramaticais. Um questionamento muito recorrente a partir do uso da técnica do Olhar Preferencial é a direcionalidade do olhar, ou seja, a indagação do que levaria os bebês a ouvirem/se atentarem por mais tempo a estímulos gramaticais ou agramaticais. Na próxima seção, discutiremos sobre a análise dos dados perceptuais dos bebês e sobre a preferência pelos estímulos novos ou conhecidos (em relação ao padrão apresentado na fase de familiarização).

#### **4.4 – Como avaliar os dados perceptuais do bebê durante a atividade experimental**

Um questionamento bastante recorrente na análise dos resultados das atividades que utilizam a técnica tanto da Escuta preferencial quanto do Olhar preferencial é em relação à preferência dos bebês pelos estímulos gramaticais ou agramaticais (em relação ao padrão familiarizado). A respeito dessa informação, Name (2012) apresenta

Roder, Bushnell e Sasseville (2000), com base em uma série de experimentos usando estímulos visuais (não linguísticos), relacionam a escolha da criança à formação de uma representação (dos estímulos) na memória. A preferência pela novidade indicaria que a representação do estímulo conhecido/familiar já estaria armazenada na memória. A ideia é que se o estímulo é congruente com informação familiar, que é ativada na memória, a atenção do bebê a esse estímulo seria inibida; já o estímulo novo estimularia a atenção do bebê que, não tendo uma representação armazenada, começaria a construí-la, daí o maior tempo dispendido nesse caso.

De acordo com essa proposta, a preferência pelo ensaio conhecido ou novo passaria por, pelo menos, três estágios no processo de construção de representações: um estágio inicial, anterior à exposição efetiva a um dado estímulo; um estágio intermediário, correspondendo à preferência pelo estímulo conhecido/familiar; e um estágio final, referente à preferência pelo não familiar/novidade, que sinalizaria um desinteresse ou inibição da atenção ao estímulo cuja representação já estaria formada na memória. O tempo de escuta maior para a condição grammatical poderia ser resultante de algo já anteriormente existente na memória da criança e o tempo de escuta maior para a condição agramatical poderia ser explicado pelo desinteresse pelo conhecido e estranhamento pelo novo. Entretanto, conforme Name (2012), essa proposta de estágios não seria efetivamente suficiente para explicar o comportamento do bebê nos experimentos de percepção linguística, pela dificuldade em se definir o estágio inicial, anterior à exposição do bebê à língua – seja ela qual for. A preferência pelo familiar poderia ser explicada justamente de modo inverso: pelo reconhecimento de algo já representado na memória da criança, como, p.ex., itens funcionais de sua língua em contraste com pseudoitens.

Dessa forma, vale ressaltar que, para que haja uma análise significativa dos resultados apresentados pelos bebês em relação ao padrão apresentado na fase de familiarização, faz-se necessário que os resultados indiquem que os bebês são capazes de distinguir os ensaios

apresentados. Ao exibir uma possível diferenciação entre os ensaios, o participante mostra ao pesquisador que não tratou todos os estímulos de forma semelhante, pois conseguiu perceber que entre eles havia diferenças. Importa-nos que o conjunto dos ensaios não seja tratado como um pacote único, igual ou homogêneo e que a percepção entre os ensaios gramaticais ou agramaticais ocorra de modo diferenciado. Se a criança irá “olhar” por mais tempo para uma ou outra condição (gramatical ou agramatical) é uma evidência que ela não está tratando os ensaios como similares e sinalizam perceber a diferença entre os dois grupos (familiarizados x novidade/ grammatical x agramatical), ou seja, entre o padrão apresentado por eles no teste e o padrão apresentado na fase de familiarização.

Uma proposta de análise sobre a preferência pelo novo/conhecido é decorrente do tipo de demanda(s) cognitiva(s) envolvida(s). Pesquisas que utilizam a técnica do Olhar/ Escuta preferencial se dividem em estudos que investigam habilidades perceptuais de reconhecimento de propriedades prosódicas, fonêmicas e/ou estruturais da língua sendo adquirida (SHADY, 1996; NAME, 2002; NAME & CORRÊA, 2003; SANTELmann & JUSCZYK, 1998; HÖHLE E WEISSENBORN, 2000; SHI e colaboradores ,2006) e estudos que envolvem outras demandas cognitivas (SAFFRAN, ASLIN E NEWPORT,1996; MARCUS e colaboradores, 1999; GÓMEZ, 2002; GERKEN, WILSON & LEWIS ,2005; HÖHLE e colaboradores, 2004; SHI e Melançon, 2010), nos quais focaliza-se a capacidade do bebê em perceber aquilo que não está aparente nos estímulos, em abstrair um padrão subjacente e generalizá-lo a novos estímulos. Segundo Name (2012), no primeiro tipo de experimentos, cabe à criança reconhecer os estímulos semelhantes à sua língua, o que geraria um tempo maior de escuta dos estímulos reconhecidos, enquanto que no segundo tipo não se trata de uma simples percepção dos dados, mas de uma percepção abstrata, mais complexa, possivelmente de maior demanda cognitiva, culminando em uma preferência pelo novo, por aquilo que destoa do padrão apresentado ou de sua língua. As atividades desenvolvidas nesta tese se encaixam no segundo tipo de experimentos, ou seja, parece exigir habilidades ligadas a níveis mais abstratos, de percepção de propriedades subjacentes (e não, aparentes) aos estímulos do que o simples reconhecimento de uma padrão, visto que o bebê precisa identificar, abstrair e generalizar DNAs e ainda utilizar informações prosódicas e categorizar itens do léxico juntamente ou a partir dessas tarefas, o que explicaria a preferência dos bebês em nossas atividades pelos estímulos novos/agramaticais.

Como vimos, a proposta de análise dos dados perceptuais dos bebês de Name (2012) é compatível com a grande maioria dos estudos na área e, particularmente, com os obtidos nesta

tese. No entanto, observamos que ainda se encontra sem explicação a preferência dos bebês canadenses pelos estímulos da condição grammatical na atividade experimental 2.

## 5- CONCLUSÃO

Buscou-se nesta tese investigar a abstração e generalização de dependências não-adjacentes e a categorização de itens da língua a partir da integração de mecanismos estatísticos e de habilidades computacionais/simbólicas para o tratamento das informações de natureza linguística.

Vimos, no capítulo 2, que dependências não adjacentes se apresentam como frequentes nas línguas naturais e são definidas como a co-ocorrência de marcadores morfossintáticos e/ou fonológicos com material interveniente. Nesta tese, dois aspectos principais foram destacados: i) DNAs serem constituídas por elementos funcionais (itens funcionais, *word markers* e base lexical interveniente) e ii) DNAs se apresentarem frequentemente em fronteiras de domínios prosódicos, como o grupo clítico e o sintagma fonológico.

Assumimos, no âmbito da teoria linguística, a vertente gerativista (Chomsky 1965 e obras posteriores), a qual postula que o bebê nasce dotado de um sistema cognitivo biologicamente especificado que lhe permite desenvolver a capacidade linguística e adquirir, de forma natural e espontânea, uma e qualquer língua natural; ainda, que o bebê possui habilidades de identificar no continuum sonoro propriedades que facilitarão a aquisição de sua língua materna. Assim, para que a criança adquira sua língua materna, ela terá que identificar, na interface fônica, o que há de regular na língua, baseando-se na frequência, regularidade, padrões prosódicos e fonéticos disponíveis no fluxo da fala (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997). A partir dessas pistas, a criança identificará os elementos funcionais da língua, os quais contêm informação acerca das propriedades gramaticalmente relevantes e são responsáveis por proverem a junção entre o léxico e a sintaxe. A criança também detectará certos padrões regulares e sistemáticos presentes nos elementos funcionais, ou seja, perceberá as variações morfonológicas nesses elementos, as quais, posteriormente, serão representadas como distinções morfonológicas. Em nosso estudo, portanto, não somente assumimos que bebês nascem com uma predisposição biológica para a aquisição de uma língua natural, mas também que o input linguístico apresenta propriedades que servirão como pistas – como a relação de DNAs entre itens funcionais, e ambientes prosódicos específicos, como fronteiras de sintagma fonológico – que facilitarão a aquisição lexical.

Como vimos no capítulo 3, vários são os estudos (SAFFRAN e colaboradores, 1996; GÓMEZ, 2002; GÓMEZ & LAKUSTA, 2004; GÓMEZ & MAYE, 2005; VAN DEN BOS et

al., 2012 e LANGUS et al., 2012) conduzidos com línguas totalmente artificiais que investigam habilidades de identificação de padrões e assumem habilidades estatísticas no tratamento das informações. Esses estudos defendem que habilidades estatísticas e probabilísticas, de domínio geral, subjazem ao tratamento de DNAs por bebês e adultos. Outros estudos (MARCUS e colaboradores, 1999; PEÑA e colaboradores 2002), assim como o nosso, concordam que mecanismos estatísticos podem ser suficientes para o reconhecimento/rastreamento de elementos que se apresentam em configurações de DA e DNA, mas defendem que, para o uso produtivo dessas configurações – abstraindo um padrão e generalizando-o em novos enunciados, a criança ou adulto precisaria de habilidades computacionais, simbólicas. De acordo com essa proposta, também defendida nesta tese, habilidades estatísticas, sozinhas, de domínio geral, não dão conta do tratamento de DNAs em língua natural e, portanto, bebês recorrem tanto a mecanismos gerais quanto a mecanismos especializados computacionais/simbólicos.

A tese defendida é de que o estabelecimento de relações entre elementos de DNAs é parte da etapa inicial de aquisição lexical e que, para tal estabelecimento, além de habilidades perceptuais, faz-se necessário o uso de habilidades computacionais – de manipulação de símbolos através de algoritmos – que torne possíveis (i) o reconhecimento e a identificação de elementos inter-relacionados, que se apresentam em configuração de DNAs, e (ii) a abstração e generalização desses padrões a novos estímulos.

Assumimos, portanto, o que chamamos de uma abordagem mista/integrada ao adotarmos a proposta do Bootstrapping Prosódico (MORGAN & DEMUTH, 1996; CHRISTOPHE et al., 1997) e uma proposta de integração mista no tratamento do estímulo linguístico (GERVAIN & MEHLER, 2010) para o processo inicial de aquisição lexical. Assim, DNAs podem ser percebidas através de mecanismos estatísticos, mas assumimos que a capacidade de abstrair padrões de DNAs e generalizá-los depende de capacidades computacionais e não meramente estatísticas.

Com o intuito de investigar as etapas iniciais da aquisição lexical, focalizando a abstração e generalização de dependências não-adjacentes e a categorização de itens das categorias Determinante (D) e Nome (N), a partir da integração de tratamentos estatísticos e de habilidades computacionais/simbólicas para o tratamento das informações linguísticas, duas atividades experimentais, apresentadas no capítulo 4, foram desenvolvidas.,

A primeira atividade experimental buscou investigar se um ambiente prosódico específico – DNAs alinhadas as fronteiras sintagma fonológico, poderia facilitar o aprendizado de DNAs em bebês canadenses e brasileiros com média de idade de 11 meses.

Esse constituinte é considerado o constituinte prosódico mais importante para a interface sintaxe-prosódia, visto que as fronteiras deste tipo de constituinte coincidem não só com fronteiras de palavras, mas também com fronteiras sintáticas (NESPOR e VOGEL, 1986). Buscamos, dessa forma, investigar se a presença de DNAs alinhadas à fronteira de sintagma fonológico seria um facilitador para a abstração e generalização do padrão de DNA estabelecido entre Det e terminação de N. Assim, a partir de dados do PB, buscamos condições em que pudéssemos alocar DNAs de duas formas distintas: alinhadas ou não a fronteiras de sintagmas fonológicos. Os ambientes escolhidos para as duas condições foram o DP e o VP. Foram criadas duas condições experimentais: Condição Adjetivo e Condição Verbo. As duas condições possuem o mesmo número de elementos (três), os mesmos “elementos” superficialmente (Det+N+Adj ou Det+N+V) e o mesmo padrão (combinação) de DNA (entre o determinante e a terminação no N). Dessa forma, a única diferença entre as duas condições é o contorno prosódico característico de um DP ou de VP. Assim, se sintagmas fonológicos representam ambiente produtivo para a interface sintaxe-prosódia, o conjunto dessas informações/pistas, ou seja o alinhamento de DNA a fronteiras de sintagmas fonológicos, deverá facilitar a abstração e generalização dos padrões das DNAs. A mesma atividade foi aplicada, utilizando a técnica do Olhar preferencial, em duas amostras distintas de participantes – bebês canadenses (1A) e bebês brasileiros (1B). Ao eleger grupos distintos de participantes, buscamos confrontar os resultados, analisando o tipo de tarefa exigida para cada grupo de participantes.

A segunda atividade experimental buscou investigar se bebês brasileiros, com média de idade de 11 meses, são capazes de categorizar linguisticamente elementos do léxico a partir do reconhecimento, abstração e generalização de DNAs. Vimos que DNAs são fundamentais para o processo de aquisição de uma língua natural e um dos fatores para sua importância deve-se ao fato de DNAs expressarem a (sub)categorização de novas palavras, ainda que o material interveniente seja bastante variável, como a subclassificação de N masculinos e femininos no português. Ao propormos essa atividade, buscamos não somente investigar a capacidade de abstração e generalização de padrões linguísticos por bebês durante o primeiro ano de vida, mas, principalmente, investigar os fatores facilitadores desse processo, assim como o uso de tais capacidades na categorização inicial de elementos do léxico. A partir do ambiente linguístico da atividade desenvolvida anteriormente (TEIXEIRA, 2102), ou seja, o DP, buscamos condições em que pudéssemos investigar a categorização dos elementos do léxico a partir do reconhecimento, abstração e generalização de DNAs, o que não foi possível de ser investigado na atividade anterior.

A partir de dados de subclassificação do léxico no PB, uma das possibilidades encontradas foi testar a forma do Det com a terminação do N, caracterizando uma relação de DNA, e avaliar em que medida bebês agrupam determinantes em uma mesma categoria a partir da identificação, abstração e generalização dessa relação de DNA. Foram utilizados determinantes masculinos do PB e pseudónomes dissílabos e paroxítonos. Utilizando a técnica do olhar preferencial, bebês brasileiros foram testados. Assim, se bebês, com média de idade de 11 meses, são capazes de categorizar os elementos do léxico a partir da identificação, abstração e generalização do padrão de DNA entre Det e terminação do N, mesmo não tendo escutado uma determinada combinação entre Det e terminação de N, agrupará os Dets em uma mesma categoria (visto que eles foram familiarizados com uma outra terminação na fase de familiarização).

Os resultados da atividade experimental 1 foram ao encontro da nossa previsão experimental inicial. Assim, bebês brasileiros e canadenses, expostos a padrões de DNAs em duas condições distintas – alinhadas a fronteiras de sintagma fonológico (Condição Verbo) e não alinhadas a fronteiras de sintagma fonológico (Condição Adjetivo) -, foram capazes de identificar, abstrair e generalizar DNAs quando estavam alinhadas a fronteiras de sintagmas fonológicos do que quando estavam distantes. Se a única diferença entre a condição Adjetivo e Condição Verbo é a presença de DNA alinhada a fronteira de sintagma fonológico, gerando contornos prosódicos distintos, podemos afirmar que o conjunto dessas informações/pistas, ou seja, o alinhamento de DNA a fronteiras de sintagmas fonológicos facilitou a abstração e generalização dos padrões das DNAs a novos estímulos a partir de um curto período de familiarização (sensibilização) – 2 minutos.

Os resultados com bebês canadenses (1A) revelaram que, mesmo diante de estímulos e contornos prosódicos totalmente diferentes de sua língua materna, bebês, aos 11 meses de idade, foram capazes de identificar a relação de DNAs, abstrair e generalizá-las a novos estímulos após 2 minutos de familiarização. Em ambas as condições, os participantes foram capazes de discriminar os ensaios gramaticais e agramaticais, escutando por mais tempo ensaios agramaticais, sugerindo que eles foram capazes de identificar, abstrair e generalizar DNAs. Entretanto, resultados estatisticamente significativos entre os ensaios só foram encontrados quando DNAs estavam alinhadas a fronteiras de sintagma fonológico (Condição Verbo), sugerindo que o conjunto de informações – DNAs e um ambiente prosódico específico, contribui para um melhor desempenho dos bebês nesta condição.

Os resultados com bebês brasileiros (1B) revelaram que esses bebês, expostos a estímulos que apresentavam determinantes e contornos prosódicos de sua língua materna, mas

em uma nova subclassificação entre combinação de Det e terminação de N, também foram capazes de identificar a relação de DNAs, abstrair e generalizá-las a novos estímulos após 2 minutos de familiarização. Como ocorreu com os bebês canadenses, em ambas as condições, os participantes brasileiros foram capazes de discriminar os ensaios gramaticais e agramaticais, escutando por mais tempo ensaios agramaticais, sugerindo que eles foram capazes de identificar, abstrair e generalizar DNAs. Entretanto, embora bebês brasileiros tenham apresentado um maior diferença entre ensaios gramaticais e agramaticais na condição Adjetivo, em comparação aos bebês canadenses, resultados estatisticamente significativos entre os ensaios só foram encontrados quando DNAs estavam alinhadas a fronteiras de sintagma fonológico (Condição Verbo), sugerindo que o conjunto de informações – DNAs e um ambiente prosódico específico –, contribui para um melhor desempenho dos bebês nesta condição.

Como ressaltamos ao longo da tese, a atividade aplicada para dois grupos de participantes diferentes parece exigir diferentes tarefas. A tarefa dos bebês canadenses nos parece difícil, visto que a pseudolíngua utilizada difere de sua língua materna. Entretanto, acreditamos que a tarefa dos bebês brasileiros é mais árdua, visto que, já conhecendo os Det do PB e a subclassificação de N, eles tiveram que, a partir de um curto período de tempo, reconhecer, abstrair e generalizar um novo padrão para Det já conhecidos, e mesmo assim apresentaram uma melhor performance em ambas as condições em relação aos bebês canadenses.

Os resultados da segunda atividade experimental foram ao encontro de nossa previsão inicial, evidenciando que dois minutos de familiarização foram suficientes para permitir a categorização dos elementos a partir da combinação entre Determinantes e terminação de Nomes. Além da evidência de categorização dos itens lexicais, assumimos que bebês brasileiros, com média de idade de 11 meses, também foram capazes de identificar, abstrair e generalizar padrões de DNAs, visto que seria “requisito”, nesta tarefa especificamente, para as habilidades iniciais de categorização. Mais uma vez, o resultado obtido com bebês brasileiros pode ser comparado com o resultado obtido, pela orientadora dessa tese, com bebês canadenses. A tarefa para os bebês canadenses era identificar, abstrair e generalizar DNAs, categorizando os elementos do léxico em uma língua totalmente diferente de sua língua materna. Para os bebês brasileiros, entretanto, identificar, abstrair e generalizar DNAs, categorizando os elementos do léxico em uma nova subcategorização parece ser uma tarefa bem mais difícil.

As atividades experimentais propostas nesta tese exigiram que os bebês não somente identificassem a relação de co-ocorrência entre um elemento e outro, mas também fossem capazes de abstrair e generalizar a relação de DNA a novos estímulos não escutados anteriormente, utilizando um conjunto de pistas disponíveis no continuum e ainda capazes de categorizá-los. Dessa forma, assumimos que habilidades estatísticas, que, como vimos, também são utilizadas em outros domínios, não dão conta do tratamento de DNAs em língua natural. No que diz respeito ao domínio linguístico, sozinhas, essas habilidades não seriam suficientes para explicar a aquisição do que estaria subjacente à estrutura da língua e o modo como as crianças abstraem e generalizam tais relações estruturais. Assim, habilidades computacionais/ simbólicas e não meramente estatísticas são utilizadas no tratamento do estímulo linguístico, tal como é proposto por Mehler e colaboradores (GERVAIN & MEHLER, 2010; ENDRESS, NESPOR & MEHLER, 2009). Os resultados obtidos a partir das atividades experimentais desenvolvidas nesta tese também vão ao encontro da proposta do uso de POMPs (ENDRESS et al. 2009) no processo de aquisição de uma língua. Dessa forma, além de habilidades perceptuais, o bebê, em fase de aquisição de uma língua, seria dotado de mecanismos – como os POMPs – que lhe permitam analisá-las de modo a identificar propriedades morfonológicas e posicionais de itens lexicais, o que contribui para a integração dos modelos simbólicos e estatísticos na mente. Ao identificar uma regularidade presente nas fronteiras de sintagma fonológicas, os bebês reconhecem padrões de identidade, permitindo a abstração e generalização de DNAs a partir de uma estrutura subjacente. De maneira complementar, ao apresentarem evidência de categorização lexical (ainda que rudimentar) a partir do tratamento de DNAs em estímulos de uma pseudolíngua, tal capacidade pressupõe, além de tratamento estatístico dos estímulos, habilidades computacionais, caracterizadas como habilidades de domínio específico, que permitem o tratamento simbólico dos dados. Assim, bebês recorrem tanto a mecanismos gerais quanto a mecanismos computacionais/simbólicos, especificamente linguísticos, na aquisição de uma língua.

O desenvolvimento da tese e os resultados obtidos suscitam a ampliação dos estudos. Investigar se existem outros ambientes/pistas facilitadores na identificação, abstração e generalização de DNAs poderia ser um início. Além disso, testar a atividade experimental 1 em crianças mais jovens que 11 meses poderia nos dar mais indícios sobre o papel da integração de múltiplas pistas, mais especificamente das DNAs em ambiente prosódico específico, no processo de aquisição lexical.

Ampliar a tarefa de categorização, expandindo-a além da categorização linguística para a categorização léxico-mundo a partir da identificação, abstração e generalização de DNAs também se apresenta como desdobramento possível desta tese.

Por fim, outra questão a ser desenvolvida concerne à sensibilidade de adultos às pistas apresentadas nas atividades, investigando se essa habilidade de identificação, abstração e generalização utilizando múltiplas pistas permanece na fase adulta, ou seja, se é algo disponível mesmo após o período de aquisição de uma língua ou se somente está disponível nesse período inicial de aquisição lexical.

## 6- REFERÊNCIAS

BAKER, M. C. **Lexical categories** - verbs, nouns and adjectives. Cambridge: Cambridge University Press, 2003.

CHOMSKY, N. **Derivation by Phase**. Working Paper, MIT., 1999.

\_\_\_\_\_. **Minimalist Inquiries: the framework**. MIT Occasional Papers in Linguistics, Number 15. MIT, 1998.

\_\_\_\_\_. **The Minimalist Program**. Mass.: The MIT Press., 1995.

\_\_\_\_\_. **Knowledge of language**. Nova York: Praeger, 1986.

\_\_\_\_\_. **Lectures on Government and Binding**. Dordrecht: Foris, 1981.

CHRISTOPHE, A.; GUASTI, T.; NESPOR, M.; DUPONUX, E.; VAN OYEN, B. **Reflections on phonological bootstrapping**: its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, v. 12, n° 5/6, p. 585-612, 1997.

\_\_\_\_\_; DUPONUX, E.; BERTONCINI, J.; MEHLER, J. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. **Journal of The Acoustical Society of America**, 95, 1994, p. 1570-1580.

CINQUE, G. **Adverbs and Functional heads: A CrossLinguistic Perspective**. New York: Oxford University Press, 1999.

\_\_\_\_\_. & RIZZI, L. The cartography of syntactic structures. **CISCL Working Papers**, Siena, v. 2, p. 42 – 58, 2008.

CORRÊA, L. M. S. Uma hipótese para a aquisição do gênero gramatical com particular referência para o português. Letras de Hoje. Porto Alegre: EDIPUCRS, 2001. Conciliando processamento linguístico e teoria de língua no estudo da aquisição da linguagem: habilidades discriminatórias de bebês, categorias funcionais e a disponibilidade de um sistema computacional linguístico. In. Corrêa, L. M. S. (org.), **Aquisição da Linguagem e Problemas do Desenvolvimento Linguístico**. Editora da PUCRJ. Rio de Janeiro, 2006.

\_\_\_\_\_. Aquisição e processamento da linguagem: uma abordagem integrada sob a ótica minimalista. **Gragoatá**, [S.l.], v. 16, n. 30, 2011.

CRYSTAL, D. **A Dictionary of Linguistics and Phonetics**. 3<sup>rd</sup> edition, Cambridge, Mass.: Blackwell, 1994.

- CUNHA, Celso & CINTRA, Lindley. **Nova Gramática do Português Contemporâneo**. 5 ed. Rio de Janeiro: Lexicon, 2008.
- ENDRESS, A.D; NESPOR, M; MEHLER, J. Perceptual and memory constraints on language acquisition. **Trends in Cognitive Sciences**. Vol.13, n8, 2009.
- \_\_\_\_\_ ; DEHAENE-LAMBERTZ G, MEHLER J. Perceptual constraints and the learnability of simple grammars. **Cognition** 105(3):577–614, 2007.
- \_\_\_\_\_ ; SCHOLL BJ, MEHLER J. The role of salience in the extraction of algebraic rules. **J. Exp. Psychol.:Gen.** 134(3):406–19, 2005.
- FISER J, ASLIN RN. Statistical learning of higher-order temporal structure from visual shape sequences. **J. Exp. Psychol.: Learn. Mem. Cogn.** 28(3):458–67, 2002.
- \_\_\_\_\_. Statistical learning of new visual feature combinations by infants. **Proc. Natl. Acad. Sci. USA** 99(24):15822–26, 2002.
- FODOR, J. A. **Modularity of Mind**. Cambridge, MA: MIT Press, 1983.
- FROTA, S. **Prosody and focus in European Portuguese. Phonological phrasing and intonation**. New York: Garland Publishing, 2000.
- GELMAN, S.A. Domain Specificity. In: WILSON, R. A.; KEIL, F.C. (eds.) **The MIT Encyclopedia of the Cognitive Sciences**. Massachusetts: Massachusetts Institute of Technology, 239, 1999.
- GERVAIN, J. & MEHLER, J. Speech Perception and Language Acquisition in the First Year of Life. **Annual Review of Psychology**. Vol.61, pp.191-218, 2010.
- GLEITMAN, L. The structural sources of verb meanings. **Language Acquisition**, v.1, p.3-55, 1990.
- \_\_\_\_\_. The structural sources of verb meaning. In: P.BLOOM (ed). **Language acquisition: Core Readings**. Cambridge, Mass.: MIT Press, p.174-221, 1994.
- GÓMEZ, R.L. Variability and detection of invariant structure. **Psychological Science**, 13, 431–436, 2002.
- \_\_\_\_\_ & LAKUSTA, L. A first step in form-based category abstraction by 12-month-old infants. **Developmental Science**, 7, 567-580, 2004.
- \_\_\_\_\_ & MAYE, J. The Developmental Trajectory of Nonadjacent Dependency Learning. **Infancy**, 7, 183- 206, 2005.
- GOUT, A.; CHRISTOPHE, A. O papel do *bootstrapping* prosódico na aquisição da sintaxe e do léxico. In: CORRÊA, L.M.S. (Org). **Aquisição da linguagem e problemas do desenvolvimento linguístico**. Rio de Janeiro: Ed. PUC-Rio; São Paulo: Loyola, 2006.
- \_\_\_\_\_ & MORGAN, J.; Phonological phrase boundaries constrain lexical Access: II Infant data. **Journal of Memory and Language**, 51, p. 548-567, 2004.
- GUASTI, M.T. Language acquisition. **The growth of grammar**, MIT press. 2002.

- GUSSENHOVEN, C. **The phonology of tone and intonation.** Cambridge: Cambridge University Press, 2004.
- HAUSER, M.; CHOMSKY, N.; FITCH, W. T. The Faculty of language: what is it, who has it, and how did it evolve? **Science**, v. 298, p.1569- 1579, 2002.
- \_\_\_\_\_ ; NEWPORT, E.L, ASLIN, R.N. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. **Cognition** 78(3):B53–64, 2001.
- HIRSH-PASEK, K.; NELSON, D. G. K.; JUSCZYK, P. W. CASSIDY, K. W.; DRUSS, B.; KENNEDY, L. Clauses are perceptual units for young infants. **Cognition**, 26, p. 269-286, 1987.
- HOHLE, B. & WEISSENBORN, J. German-learning infants' ability to detect unstressed closed-class elements in continuous speech. **Developmental Science** 6, 122-127, 2003.
- HORST, S. Computational Theory of Mind. In: **The MIT Encyclopedia of the Cognitive Sciences.** Massachusetts: Massachusetts Institute of Technology, 170, 1999.
- JUSCZYK, P. W.; CUTLER, A.; REDANZ, N. Infants' Preference for the Predominant Stress Patterns of English Words. **Child Development**, v. 64, p. 675-687, 1993.
- \_\_\_\_\_ ; KEMLER-NELSON, D. G.; HIRSH-PASEK, K.; KENNEDY, L.; WOODWARD, A.; PIWOZ, J. Perception of acoustic correlates of major phrasal units by young infants. **Cognitive Psychology**, 24, p. 252-293, 1992.
- KEMLER NELSON, D. G.; HIRSH-PASEK, K.; JUSCZYK, P. W.; CASSIDY, K. W. How the prosodic cues in motherese might assist language learning. **Journal of Child Language**, 16, p. 55-58, 1989.
- KUDO N; NONAKA Y; MIZUNO K; OKANOYA K. **Statistical learning and word segmentation in neonates:** an ERP evidence. Presented at Annu. Meet. XVth Bienn. Int. Conf. Infant Studies, Kyoto, Japan, 2006.
- LANGUS, L; MARCHETTO, E; BION, R.A.H; NESPOR, M. Can prosody be used to discover hierarchical structure in continuous speech? **Journal of Memory and Language**, 66, p. 285- 306, 2012.
- LEWIS, R. L. Cognitive Modeling, Symbolic. In: **The MIT Encyclopedia of the Cognitive Sciences.** Massachusetts: Massachusetts Institute of Technology, 141-142, 1999.
- LINDENAU J. R, GUIMARÃES L.S.P. Calculando o tamanho de efeito no SPSS. **Revista HCPA.** 32(3):363-81, 2012.
- MARCUS, G. F.; VIJAVAN, S.; BANDI RAO, S.; VISHTON, P.M. Rule learning by seven-month-old infants. **Science**, 283, 77–80, 1999.
- MARLER, P. The instinct to learn. In S. Carey and R. Gelman, Eds., **The Epigenesis of Mind: Essays on Biology and Cognition.** Hillsdale, NJ: Erlbaum, 1991.

MAYER, R. E. Understanding conceptual change: A commentary. In M. Limon, & L. Mason (Eds.), **Reconsidering conceptual change: Issues in theory and practice**. Netherlands: Kluwer Academic Publishers, 2002.

MEDIN, D.L; AGUILAR, C. Categorization. In: **The MIT Encyclopedia of the Cognitive Sciences**. Massachusetts: Massachusetts Institute of Technology, 104, 1999.

MINTZ, T. H., NEWPORT, E. L., & BEVER, T. G. The distributional structure of grammatical categories in speech to young children. **Cognitive Science**, 26, 393–424, 2002.

MIOTO, C. S.; SILVA, M. C. F.; LOPES, R. E. V. **Manual de sintaxe**. 2 ed. Florianópolis: Insular, 2005.

MORAES, J. A. **Os fenômenos suprasgmentais do Português do Brasil**. Tese de Doutorado em Letras. PUC-RS, 1998.

MORGAN, J. L.; DEMUTH, K. Signal to Syntax: an overview. In: J. L. MORGAN & K. DEMUTH (Orgs.), **Signal to Syntax: Bootstrapping from speech to grammar in early acquisition**. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1-22, 1996.

\_\_\_\_\_ ; SHI, R.; ALLOPENNA, P., Perceptual Bases of Rudimentary Grammatical Categories: Toward a Broader Conceptualization of Bootstrapping. In: J. L. MORGAN & K. DEMUTH (Orgs.), **Signal to Syntax: Bootstrapping from speech to grammar in early acquisition**. Mahwah, New Jersey: Lawrence Erlbaum Associates, 263-286, 1996.

MYERS, J.; JUSCZYK, P. W.; KEMLER-NELSON, D. G.; CHARLES-LUCE, J.; WOODWARD, A. L.; HIRSH-PASEK, K. Infants' sensitivity to word boundaries in fluent speech. **Journal of Child Language**, 23(1), p. 1-30, 1996.

NAME, M.C. O que nos dizem os resultados experimentais sobre a percepção da fala pelo bebê. **Veredas** (UFJF. Online), v. Especial, p. 284-297, 2012.

\_\_\_\_\_. Bootstrapping sintático: o papel da ordem estrutural na aquisição de nomes e adjetivos. **Letras de Hoje**, v. 42, p. 53-63, 2007.

\_\_\_\_\_. **Habilidades perceptuais e linguísticas no processo de aquisição do sistema de gênero no português**. Tese (Doutorado) PUC RJ, 2002.

\_\_\_\_\_ & CORRÊA, L. M. S. Explorando a escuta, o olhar e o processamento sintático. In: In : CORRÊA, L.M.S. (org.). **Aquisição da Linguagem e Problemas do Desenvolvimento Lingüístico**. Rio de Janeiro: Editora da PUC-RJ, p. 79-100, 2006.

\_\_\_\_\_ & CORRÊA, L.M.S. Delimitação Percentual de uma Classe Correspondente à Categoria Funcional D: Evidencias da Aquisição do Português. **Fórum Linguístico**, (3)1, 55-88, 2003.

\_\_\_\_\_ & LAGUARDIA, M. C. C. T. Etapas iniciais da aquisição lexical. **Prolíngua** (João Pessoa), v. 8, p. 296-308, 2013.

\_\_\_\_\_; SHI, R.; KOULAGUINA, E.; Non- adjacent dependency learning and grammatical categorization in 11- month-old infants. **12th International Congress for the Study of Child Language.** Montreal, 2011.

NESPOR, M.; VOGEL, I. **Prosodic Phonology:** with a new foreword. Dordrecht-Holland: Foris Publications, 2007.

NEWPORT, E.L.; ASLIN, R.N. Learning at a distance: I. Statistical learning of non-adjacent dependencies. **Cognitive Psychology**, 48, 127-162, 2004.

PEÑA, M., BONATTI, L., NESPOR, M., & MEHLER, J. Signal-driven computations in speech processing. *Science*, 298, 604-607, 2002.

PINKER, S. **The Language Instinct.** New York: Penguin Books, 1994

\_\_\_\_\_. **Language learnability and language development.** Cambridge, MA: Harvard University Press, 1984.

REDINGTON, M., CHATER, N., & FINCH, S. Distributional information: A powerful cue for acquiring syntactic categories. **Cognitive Science**, 22, 425–469, 1998.

SAFFRAN, J. R.; ASLIN, R.; NEWPORT, E. L. Statistical learning by 8-month-old infants. **Science**, 274, p. 1926-1928, 1996.

\_\_\_\_\_.; NEWPORT, E. L.; ASLIN, R. Word smentation: The role of distributional cues. **Journal of Memory and Language**, 35, p.606-621, 1996.

\_\_\_\_\_. & THIESSEN, E.D. Domain general learning capacities. In *E. Hoff and M. Schatz (Eds.) Blackwell Handbook of Language Development*, pp. 68-86. Oxford: Blackwell Publishing, 2007

SANTELmann, L.; JUSCZYK, P. **Sensitivity to discontinuous dependencies in language learners:** evidence for limitations in processing space. **Cognition**, 69, 105-134, 1998.

SCHAFER, G.; PLUNKETT, K. & HARRIS, P.L. Rapid word learning by 15-month-olds under tightly controlled conditions. **Child Development**, 69; p.309-320, 1998.

SHADY, M. **Infants' sensitivity to function morphemes.** New York: State University of New York at Buffalo, 1996.

SMITH, L.B. How domain-general process may create domain-specific biases. In: BOWERMAN, M and LEVINSON, S. (Orgs) **Language acquisition and conceptual development.** Cambridge: Cambridge University Press. 101-131, 2001.

SPELKE, E. S. Initial knowledge: Six suggestions. **Cognition** 50: 431– 445, 1994.

TEIXEIRA, M. C. C. **A contribuição das informações distribucionais de DNA nas etapas iniciais da aquisição lexical.** Dissertação de Mestrado. Universidade Federal de Juiz de Fora, 2012.

\_\_\_\_\_. & NAME, M. C. Abstração e generalização de dependências não adjacentes entre Determinantes e Nomes nas etapas iniciais do processo de aquisição lexical. In: **VIII ENAL / II EIAL Encontro Inter/Nacional sobre Aquisição da Linguagem**, 2012,

Juiz de Fora. Anais do VIII ENAL / II EIAL Encontro Inter/Nacional sobre Aquisição da Linguagem. Juiz de Fora. v. 1. p. 61-70, 2011.

TORO J.M, TROBALON J.B. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67(5):867–75, 2005.

VAN DEN BOS, E., CHRISTIANSEN, M. H., & MISYAK. Statistical learning of probabilistic nonadjacent dependencies in an artificial language: multiple cues make it possible. **Lorentz Center Workshop**, Leiden, the Netherlands, 2012.

VAN HEUGTEN, M. & SHI, R. Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *J. Acoust. Soc. Am.*, 128(5), 223-228, 2010.

## Glossário

### Abstração e generalização

Por abstração e generalização<sup>27</sup> entende-se que são habilidades de reconhecimento de novas sequências como exemplares de um padrão aprendido. Assim, a criança pode nunca antes ter ouvido uma sequência específica da língua, mas ser capaz de reconhecê-la como um exemplo do padrão linguístico aprendido. Em nossas atividades experimentais, se uma criança for sensível aos elementos a ela apresentados, for capaz de aprender o padrão da pseudolíngua a qual ficou exposta durante um curto período de tempo, identificando as relações de DNAs, ela será capaz de abstrair aquele padrão e generalizá-lo, no sentido de gerar uma regra algébrica que será aplicada a novas estruturas não ouvidas anteriormente.

### Aprendizagem estatística

Aprendizagem estatística é a capacidade que os seres humanos e outros animais possuem que lhes permitem extrair regularidades estatísticas do mundo em torno deles para aprender sobre o meio. Embora a aprendizagem estatística seja utilizada, na atualidade, para o aprendizado de forma generalizada, o fenômeno foi identificado pela primeira vez na aquisição da linguagem humana em estudo realizado por Saffran, J; Aslin, R & Newport, E (1996), como veremos no decorrer desta tese, sugerindo que as crianças são capazes de identificar relações estatísticas entre sílabas mesmo com uma exposição muito limitada a um estímulo.

Na última década, a aprendizagem estatística (*SL- Statistical Learning*) tem emergido especialmente como um mecanismo chave proposto para a aquisição de dependências probabilísticas inerentes no sinal do fluxo de fala, dessa forma

---

<sup>27</sup> Não há na literatura definições bem estabelecidas desses termos. Além disso, de acordo com Jenny Saffran, em conversas trocadas por e-mail, os dois termos têm sido usados de modo intercambiável. Neste trabalho, entretanto, optamos por usar os dois termos por acreditar que, ainda que intercambiáveis na literatura, a generalização de um padrão parece exigir que a abstração do mesmo tenha ocorrido.

Aprendizagem estatística é um fenômeno intimamente temporal envolvendo a codificação de regularidades sequenciais desdobradas ao longo do tempo e do espaço e na formação simultânea de conhecimentos distribucionais através da aprendizagem contínua  
 (Misyak; Christiansen; Tomblin, 2009)

Como veremos na seção 3 desta tese, embora o fenômeno da aprendizagem estatística tenha sido descoberto pela primeira vez no contexto da aquisição da linguagem e suas descobertas, ainda hoje, se concentram nesta área, várias são as evidências de que esse tipo de aprendizagem seja uma habilidade de domínio geral e, muito provavelmente, não exclusivo para os seres humanos (Turk-Browne et al, 2005). Em domínios não-auditivos, há evidências de que os seres humanos são capazes de aprender informação visual estatística se essa informação for apresentada ao longo do espaço, por exemplo (FISER & ASLIN, 2001) ou ao longo do tempo (FISER & ASLIN, 2002).

Evidências da aprendizagem estatística também foram encontradas em outros primatas (NEWPORT ET AL. 2004) e capacidades de aprendizagem estatísticos limitadas também foram encontradas em não-primatas, como em ratos (TORO & TROBALÓN, 2005). Como veremos, no capítulo 3 desta tese, ainda que esses resultados possam sugerir que a aprendizagem estatística possa ser um mecanismo de aprendizagem geral, utilizado na aquisição da linguagem, essa habilidade sozinha, parece não ser suficiente para o tratamento de DNAs em língua natural e, portanto, bebês recorreriam tanto a mecanismos gerais quanto a mecanismos computacionais/simbólicos, especificamente linguísticos, na aquisição de uma língua.

## **Aquisição lexical**

Este trabalho abrange a primeira etapa do processo de aquisição lexical. Nesta etapa, explorada neste trabalho com média de idade de 11 meses, o bebê balbucia poucas palavras, mas revela ter um conhecimento sobre a língua muito maior do que o demonstrado. Nessa fase, a criança, exposta ao estímulo linguístico de sua língua e dotada de habilidades perceptuais e computacionais, necessita fragmentar o *continuum* linguístico ao qual está exposta em unidades menores. Tal tarefa não parece simples, visto que o estímulo linguístico é um emaranhado de sons, muitas vezes sem ao menos pausas definidas. Assim, antes de

identificar as unidades de palavras, os bebês segmentam o fluxo da fala em unidades cada vez menores e relevantes para o processamento linguístico e essa segmentação proporciona o reconhecimento dos padrões fonotáticos, distribucionais e prosódicos de sequências fonológicas, os quais servirão como pistas para a aquisição lexical e a diferenciação entre itens lexicais e funcionais, como veremos no modelo de *bootstrapping* prosódico. Cabe ressaltar que esses processos aconteceriam mais ou menos juntos – de segmentação, de reconhecimento dos padrões fonotáticos, distribucionais, dos parâmetros prosódicos, “se alimentando” uns aos outros.

### Categorização

“É o processo pelo qual entidades distintas são tratadas como equivalentes e é uma das atividades cognitivas mais fundamentais, pois nos permite compreender e fazer previsões sobre objetos e eventos do mundo” (MEDIN; AGUILAR, 1999, p. 104). Em se tratando de categorização linguística, ao percebermos que dois ou mais elementos possuem algumas características em comum, assim como ocorreu em nossa atividade experimental 2, por compartilharem as mesmas características, serão agrupados em uma mesma categoria. Em nossa atividade, e de maneira geral nesta tese, estamos utilizando a expressão “categorização inicial” por investigarmos somente a categorização dos itens (por exemplo, apesar de não ter ouvido a combinação “**Nosso** e pebe”- Nosso+X-e -, a criança será capaz de agrupar o determinante nosso com o determinante seu, já que anteriormente o determinante “Seu” havia sido combinado com pebe – “**Seu** pebe”) e não a categorização e mapeamento entre léxico e mundo.

### Habilidades Computacionais/simbólicas

“Segundo a teoria Computacional da mente (CTM - *Computational Theory Of Mind*), a mente funciona como um computador digital: um dispositivo discreto que armazena representações simbólicas e as manipula de acordo com as regras sintáticas” (HORST & STEVEN, 1999). O fundamento teórico do modelo cognitivo é a ideia de que a cognição é um tipo de computação. A ideia principal é de que o que a mente faz, em parte, é realizar tarefas cognitivas através da computação. Segundo Lewis (1999), um modelo cognitivo é considerado um modelo cognitivo simbólico se possui propriedades de um sistema simbólico,

como na hipótese do sistema simbólico físico (PSSH) de Newell and Simon's (Newell and Simon's, 1976 *apud Lewis*, 1999). O PSSH defende a hipótese de condições necessárias e suficientes para um sistema físico realizar a inteligência. Trata-se de uma reformulação da máquina de *Turing* - modo conceptual de dizer que qualquer função bem definida poderia ser executada, passo a passo, de acordo com a simples regra : “se você está em um estado P e possui o *input* Q, então faça R”-, que identifica o processamento simbólico como requisito para a cognição complexa. Assim, como o cérebro é um dispositivo o qual *input* e *output* podem ser caracterizados em termos de funções matemáticas , em sentido bastante abstrato, poderia ser simulado pela máquina de *Turing*. A condição é que o sistema seja capaz de manipular e compor símbolos e estruturas simbólicas – padrões físicos com processos associados que fornecem ao padrão o poder de denotar tanto entidades externas quanto estruturas simbólicas internas.

Diferentemente, Fodor (1983) concebe sua hipótese da modularidade da mente assumindo uma concepção de mente computacional, entretanto defende que apesar de serem essencialmente associáveis (visto que ambos são mecanismos de manipulação de símbolos), a arquitetura mental não pode ser entendida como uma “máquina de *Turing*”. Nossa proposta se assemelha ao que é defendido por Fodor. Segundo Fodor, diferentemente de um computador ou “máquina de *Turing*”, para a computação mental a troca de informação com o ambiente é fundamental, enquanto que para o computador, o que é externo é irrelevante para seu desempenho. Esta proposta torna-se muito relevante para a aquisição lexical, visto que para a efetivação desse processo torna-se necessário a disponibilidade de mecanismos capazes de captar a informação do estímulo linguístico (informação externa) e fazê-los acessíveis aos sistemas computacionais. Dessa forma em sua Hipótese da modularidade da mente (1983), Fodor defende que a arquitetura mental necessita de i) um sistema cognitivo cujas operações cruzem domínios; ii) mecanismos computacionais que gerem hipóteses sobre um dado objeto a partir da representação imediata deste objeto (sistema de *input* ou módulos) e iii) mecanismos que captem informação externa e a disponibilizem em um formato legível aos sistemas computacionais.

De acordo com Churchland and Grush (1999) , há duas visões principais que se referem a questão de como, de fato, o cérebro representa e computa. A primeira explora o modelo serial familiar, computação digital, nos quais representações são símbolos, assim como sentenças são símbolos e computações são regras formais (algoritmos) que operam sob os símbolos, como por exemplo regras do tipo “ Se então” podem se estender em forma lógica

e design circuito. A segunda visão está ancorada na neurociência, investigando como as células do cérebro (neurônios) respondem a sinais externos como luz e som, como integrar esses sinais para extrair informações de ordens superiores e como, em um estágio posterior, neurônios interagem na produção de decisões e comandos motores. Nossa trabalho se insere na primeira visão, visto que buscamos investigar as habilidades computacionais/simbólicas, especificamente na abstração de regras algébricas. Como vemos em nossas atividades se um elemento X é combinado com um elemento Y, mesmo se os elementos intervenientes variarem, a regra ainda é a mesma, pois foi generalizada a novas instâncias. Da mesma forma, se elementos X e Y combinam com elemento Z, mesmo não tendo escutado a combinação Y e W, ela será considerada de acordo com a regra abstraída, pois X e W obedecem à regra e X e Y são postos em uma mesma categoria.

## Anexos

### Anexo 1 – Estímulos utilizados na 1<sup>a</sup> atividade experimental – Condição Adjetivo e Condição Verbo

Condição Adjetivo	
Grupo 1 : [Nossa/Essa+N-o+ Adj-il] φ e [Meu/Seu+N-e+Adj-il] φ	Grupo 2: [Meu/Seu+N-o+ Adj-il] φ e [Nossa/Essa+N-e+Adj-il] φ
<b>Familiarização</b> Nossa vato pigil Seu tuge bilil Essa dabo munil Meu dale cibil Nossa rofo tudil Seu lepe gidil Essa mebo tebil Meu vebe napil Nossa fovo ceril Seu pile vupil Essa pifo melil Meu fide dunil Essa vato mulil Meu tuge dipil Nossa dabo diliu Seu dale rulil Essa rofo sinil Meu lepe tunil Nossa mebo dalil Seu vebe cutil Essa fovo defil Meu pile lutil Nossa pifo pinil Seu fide befil	<b>Familiarização</b> Nossa tuge pigil Seu vato bilil Essa dale munil Meu dabo cibil Nossa lepe tudil Seu rofo gidil Essa vebe tebil Meu mebo napil Nossa pile ceril Seu fovo vupil Essa fide melil Meu pifo dunil Essa tuge mulil Meu vato dipil Nossa dale diliu Seu dabo rulil Essa lepe sinil Meu rofo tunil Nossa vebe dalil Seu mebo cutil Essa pile defil Meu fovo lutil Nossa fide pinil Seu pifo befil
Os participantes ouviram a lista 2 vezes seguidas	Os participantes ouviram a lista 2 vezes seguidas

**Ensaios do teste**

<b>Ensaio 1</b>	<b>Ensaio 2</b>	<b>Ensaio 3</b>	<b>Ensaio 4</b>	<b>Ensaio 5</b>	<b>Ensaio 6</b>
Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.
Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.
Essa bifo duvil.	Meu leto duvil.	Essa bifo duvil.	Meu leto duvil.	Essa bifo duvil.	Meu leto duvil.
Seu mape duvil.	Nossa rufe duvil.	Seu mape duvil.	Nossa rufe duvil.	Seu mape duvil.	Nossa rufe duvil.
Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.
Essa bifo duvil.	Nossa rufe duvil.	Essa bifo duvil.	Nossa rufe duvil.	Essa bifo duvil.	Nossa rufe duvil.
Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.
Seu mape duvil.	Meu leto duvil.	Seu mape duvil.	Meu leto duvil.	Seu mape duvil.	Meu leto duvil.
<b>Ensaio 7</b>	<b>Ensaio 8</b>	<b>Ensaio 9</b>	<b>Ensaio 10</b>	<b>Ensaio 11</b>	<b>Ensaio 12</b>
Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.
Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.
Essa bifo duvil.	Meu leto duvil.	Essa bifo duvil.	Meu leto duvil.	Essa bifo duvil.	Meu leto duvil.
Seu mape duvil.	Nossa rufe duvil.	Seu mape duvil.	Nossa rufe duvil.	Seu mape duvil.	Nossa rufe duvil.
Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.	Nossa bifo duvil.	Essa rufe duvil.
Essa bifo duvil.	Nossa rufe duvil.	Essa bifo duvil.	Nossa rufe duvil.	Essa bifo duvil.	Nossa rufe duvil.
Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.	Meu mape duvil.	Seu leto duvil.
Seu mape duvil.	Meu leto duvil.	Seu mape duvil.	Meu leto duvil.	Seu mape duvil.	Meu leto duvil.

<b>Condição Verbo</b>	
<b>Grupo 1 : [Nossa/Essa+N-o]ɸ [V-iu] e [Meu/Seu+N-e] ɸ [V-iu]</b>	<b>Grupo 2: [Meu/Seu+N-o] ɸ [V-iu] e [Nossa/Essa+N-e] ɸ [V-iu]</b>
<b>Familiarização</b> Nossa vato pigiu Seu tuge biliu Essa dabo muniu Meu dale cibiu Nossa rofo tudiu Seu lepe gidiu Essa mebo tebiu Meu vebe napiu Nossa fovo ceriu Seu pile vupiu Essa pifo meliu Meu fide duniu Essa vato muliu Meu tuge dipiu Nossa dabo diliu Seu dale ruliu Essa rofo siniu Meu lepe tuniu Nossa mebo daliu Seu vebe cutiu Essa fovo defiu Meu pile lutiu Nossa pifo piniu Seu fide befiu	<b>Familiarização</b> Nossa tuge pigiu Seu vato biliu Essa dale muniu Meu dabo cibiu Nossa lepe tudiu Seu rofo gidiu Essa vebe tebiu Meu mebo napiu Nossa pile ceriu Seu fovo vupiu Essa fide meliu Meu pifo duniu Essa tuge muliu Meu vato dipiu Nossa dale diliu Seu dabo ruliu Essa lepe siniu Meu rofo tuniu Nossa vebe daliu Seu mebo cutiu Essa pile defiu Meu fovo lutiu Nossa fide piniu Seu pifo befiu
Os participantes ouviram a lista 2 vezes seguidas	Os participantes ouviram a lista 2 vezes seguidas

**Ensaios do teste**

<b>Ensaio 1</b>	<b>Ensaio 2</b>	<b>Ensaio 3</b>	<b>Ensaio 4</b>	<b>Ensaio 5</b>	<b>Ensaio 6</b>
Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.
Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.
Essa bifo duviu.	Meu leto duviu.	Essa bifo duviu.	Meu leto duviu.	Essa bifo duviu.	Meu leto duviu.
Seu mape duviu.	Nossa rufe duviu.	Seu mape duviu.	Nossa rufe duviu.	Seu mape duviu.	Nossa rufe duviu.
Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.
Essa bifo duviu.	Nossa rufe duviu.	Essa bifo duviu.	Nossa rufe duviu.	Essa bifo duviu.	Nossa rufe duviu.
Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.
Seu mape duviu.	Meu leto duviu.	Seu mape duviu.	Meu leto duviu.	Seu mape duviu.	Meu leto duviu.
<b>Ensaio 7</b>	<b>Ensaio 8</b>	<b>Ensaio 9</b>	<b>Ensaio 10</b>	<b>Ensaio 11</b>	<b>Ensaio 12</b>
Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.
Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.
Essa bifo duviu.	Meu leto duviu.	Essa bifo duviu.	Meu leto duviu.	Essa bifo duviu.	Meu leto duviu.
Seu mape duviu.	Nossa rufe duviu.	Seu mape duviu.	Nossa rufe duviu.	Seu mape duviu.	Nossa rufe duviu.
Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.	Nossa bifo duviu.	Essa rufe duviu.
Essa bifo duviu.	Nossa rufe duviu.	Essa bifo duviu.	Nossa rufe duviu.	Essa bifo duviu.	Nossa rufe duviu.
Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.	Meu mape duviu.	Seu leto duviu.
Seu mape duviu.	Meu leto duviu.	Seu mape duviu.	Meu leto duviu.	Seu mape duviu.	Meu leto duviu.

**Anexo 2- Tempo individual da atividade experimental DNA com informação prosódica – bebês canadenses**

Condição Adjetivo			Condição Verbo		
GRUPO 1	GRAMATICAL	AGRAMATICAL	Grupo 1	GRAMATICAL	AGRAMATICAL
Bebê 1	4,266666667	5,55	Bebê 1	5,8	3,916666667
Bebê 2	5,816666667	3,733333	Bebê 2	4,633333333	8,633333333
Bebê 3	8,95	4,983333	Bebê 3	7,716666667	8,15
Bebê 4	8,616666667	10,28333	Bebê 4	9	8,483333333
Bebê 5	5,266666667	8,383333	Bebê 5	3,683333333	8,45
Bebê 6	5,55	5,166667	Bebê 6	7,683333333	7,183333333
Bebê 7	4,216666667	11,31667	Bebê 7	6,233333333	9,45
Bebê 8	7,816666667	8,916667	Bebê 8	4,45	7,2
GRUPO 2			GRUPO 2		
Bebê 9	14,25	11,11667	Bebê 9	5,2	6,666666667
Bebê 10	10,2	12,43333	Bebê 10	4,083333333	6,216666667
Bebê 11	7,366666667	9,283333	Bebê 11	9,916666667	11,13333333
Bebê 12	7,416666667	4,75	Bebê 12	2,883333333	4,1
Bebê 13	3,25	3,016667	Bebê 13	3,683333333	5,15
Bebê 14	9,666666667	10,43333	Bebê 14	9,866666667	7,966666667
Bebê 15	5,75	3,4	Bebê 15	3,25	2,833333333
Bebê 16	7,1	6,916667	Bebê 16	7,816666667	8,1
MÉDIA	<b>7,21875</b>	<b>7,480208</b>	Média	<b>5,99375</b>	<b>7,102083333</b>

### Anexo 3– Termo de consentimento livre e esclarecido

#### ***TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO***

Seu/sua filho(a) está sendo convidado(a) como voluntário(a) a participar da pesquisa “Etapas iniciais da aquisição lexical”. Neste estudo pretendemos observar o modo como os bebês adquirindo o português começam a smentar a fala e a reconhecer palavras. O motivo que nos leva a estudar esse assunto é contribuir para o entendimento do processo de aquisição de vocabulário desde suas etapas iniciais.

Para este estudo adotaremos o(s) suinte(s) procedimento(s): o bebê participará de uma atividade lúdica (uma “brincadeira”), durante a qual lhe apresentaremos imagens na tela do computador acompanhadas de frases curtas. Observaremos sua atenção e interesse aos estímulos apresentados. **A atividade não tem nenhum caráter de avaliação do desempenho e/ou de conhecimento da língua.** Seu único objetivo é observar o modo como a criança se relaciona com a língua em uma situação que simula uma atividade espontânea. A atividade dura cerca de 15 minutos e no total (desde a chegada do bebê, sua adaptação ao ambiente e saída) não ultrapassa 30 minutos.

Para seu/sua filho(a) participar deste estudo você não terá nenhum custo, nem receberá qualquer vantagem financeira. Você será esclarecido(a) sobre o estudo em qualquer aspecto que desejar e estará livre para participar ou recusar-se a participar. Poderá retirar seu consentimento ou interromper a participação a qualquer momento. A participação de seu/sua filho(a) é voluntária e a recusa em participar não acarretará qualquer penalidade ou modificação na forma em que é atendido(a) pelo pesquisador. O pesquisador irá tratar a sua identidade e a de seu/sua filho(a) com padrões profissionais de sigilo. Ele(a) não será identificado(a) em nenhuma publicação que possa resultar deste estudo. Este estudo apresenta risco mínimo, isto é, o mesmo risco existente em atividades rotineiras como conversar, tomar banho, ler, etc. Apesar disso, você tem assurado o direito a resarcimento ou indenização no caso de quaisquer danos eventualmente produzidos pela pesquisa.

Os resultados da pesquisa estarão à sua disposição quando finalizada. O nome de seu/sua filho(a) ou o material que indique a participação dele(a) não será liberado sem a sua permissão. Os dados einstrumentos utilizados na pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 anos, e após esse tempo serão destruídos. Este termo de consentimento encontra-se impresso em duas vias, sendo que uma cópia será arquivada pelo pesquisador responsável, e a outra será fornecida a você.

Eu, \_\_\_\_\_, portador(a) do documento de Identidade \_\_\_\_\_, fui informado(a) dos objetivos do presente estudo de maneira clara e detalhada e esclareci minhas dúvidas. Sei que a qualquer momento poderei solicitar novas informações e modificar minha decisão de autorizar meu/minha filho(a) participar se assim o desejar. Declaro que concordo em participar desse estudo. Recebi uma cópia deste termo de consentimento livre e esclarecido e me foi dada a oportunidade de ler e esclarecer as minhas dúvidas.

Juiz de Fora, \_\_\_\_ de \_\_\_\_\_ de 20\_\_\_\_ .

\_\_\_\_\_  
Assinatura do(a) participante

\_\_\_\_\_  
Assinatura do(a) pesquisador(a)

Em caso de dúvidas com respeito aos aspectos éticos deste estudo, você poderá consultar:

CEP- COMITÊ DE ÉTICA EM PESQUISA - UFJF  
PRÓ-REITORIA DE PESQUISA / CAMPUS UNIVERSITÁRIO DA UFJF

JUIZ DE FORA (MG) - CEP: 36036-900

FONE: (32) 2102-3788 / E-MAIL: cep.propesq@ufjf.edu.br

PESQUISADOR(A) RESPONSÁVEL: MARIA CRISTINA LOBO NAME

ENDEREÇO: FACULDADE DE LETRAS – UFJF CAMPUS UNIVERSITARIO -  
MARTELOS

JUIZ DE FORA (MG) - CEP: 36036-300

FONE: (32) 2102.3150 / E-MAIL: CRISTINA.NAME@UFJF.EDU.BR

## Anexo 4 – Cadastro de crianças NEALP

**NEALP** Universidade Federal de Juiz de Fora



### Cadastro de Crianças

#### Projeto em Aquisição da Linguagem

Nome da Criança: \_\_\_\_\_

Data de nascimento da Criança: \_\_\_/\_\_\_/\_\_\_

Sexo: ( ) masculino

( ) feminino

Frequenta Creche ou escola? Qual? \_\_\_\_\_

Nome do responsável: \_\_\_\_\_

Profissão do pai/mãe: \_\_\_\_\_

Telefone para contato: \_\_\_\_\_ Celular: \_\_\_\_\_

E-mail: \_\_\_\_\_

Os pais da criança são falantes nativos do Português do Brasil? Sim ( ) Não ( )

A criança viveu no exterior ou tem contato regular com falante de língua estrangeira (diferente do pai ou da mãe):      Sim ( )      Não ( )

A criança possui parentes próximos ou antepassados com qualquer problema de linguagem:  
Sim ( )      Não ( )

Marque caso a criança tenha apresentado qualquer dos quadros abaixo:

( ) otite de repetição

( ) lesão cerebral

( ) qualquer deficiência genética

Este formulário é para facilitar o contato entre o NEALP e os pais interessados em nossa pesquisa. O preenchimento do mesmo NÃO IMPLICA qualquer engajamento da parte dos pais. Toda informação nele contida é confidencial e para uso exclusivo do NEALP.

**Anexo 5- Tempo individual da atividade experimental DNA com informação prosódica – bebês brasileiros**

Condição Adjetivo			Condição Verbo		
GRUPO 1	GRAMATICAL	AGRAMATICAL	GRUPO 1	GRAMATICAL	AGRAMATICAL
Bebê 1	5,416666667	3,066666667	Bebê 1	8,05	14,43333333
Bebê 2	2,833333333	7,366666667	Bebê 2	9,283333333	10,13333333
Bebê 3	6,05	3,666666667	Bebê 3	6,85	8,3
Bebê 4	3,733333333	4,066666667	Bebê 4	4,516666667	5,566666667
Bebê 5	9,666666667	10,55	Bebê 5	5,583333333	9,016666667
Bebê 6	10,35	12,1	Bebê 6	5,833333333	11,58333333
Bebê 7	6,4	8,616666667	Bebê 7	3,066666667	4,816666667
Bebê 8	4,933333333	8,15	Bebê 8	4,966666667	7,116666667
GRUPO 2			GRUPO 2		
Bebê 9	3,65	9,633333333	Bebê 9	9,8	7,633333333
Bebê 10	7,683333333	5,25	Bebê 10	4,366666667	6,666666667
Bebê 11	7,55	7,766666667	Bebê 11	1,966666667	6,416666667
Bebê 12	4,166666667	3,2	Bebê 12	8,283333333	8,6
Bebê 13	4,183333333	3,866666667	Bebê 13	8,183333333	9,933333333
Bebê 14	4,083333333	4,783333333	Bebê 14	5,933333333	5,733333333
Bebê 15	12,38333333	11,28333333	Bebê 15	7,483333333	8,2
Bebê 16	9,266666667	10,53333333	Bebê 16	8,333333333	9,683333333
<b>MÉDIA</b>	<b>6,396875</b>	<b>7,11875</b>	<b>MÉDIA</b>	<b>6,40625</b>	<b>8,364583</b>

## Anexo 6- Estímulos utilizados na 2<sup>a</sup> atividade experimental

Experimento categorização	
Grupo 1 seu +X-e; seu/nosso + X-u; meu +X-a; meu/esse + X-E	Grupo 2 seu+X-a; seu/nosso + X-E; meu+X-e; meu/esse + X- u
<b>Familiarização</b> Seu bafe Meu Goca Nosso Chivu Esse Sogué Meu Nôca Seu Guque Nosso Tadu Meu Pilé Seu Jomu Meu Bapa Esse Maké Seu Quive Nosso Bechu Meu Puva Meu Jufé Seu Tadu Seu Zobe Esse pilé Seu Bechu Meu Tafa Seu Pôfe Meu Maké Meu Vêba Seu Mepe Esse Jufé Meu Ziba Seu dabe Nosso Jomu Meu Mefa Seu Chivú Meu Sogué Seu Nêque Nosso Tadu Esse Pilé Meu Puva Seu Jomu Seu Zobe Meu Maké Meu Bapa Seu Nêque	<b>Familiarização</b> Meu bafe Seu Goca Esse Chivu Nosso Sogué Esse Chivu Seu Nôca Meu Guque Esse Tadu Seu Pilé Meu Jomu Seu Bapa Nosso Maké Meu Quive Esse Bechu Seu Puva Seu Jufé Meu Tadu Meu Zobe Nosso pilé Meu Bechu Seu Tafa Meu Pôfe Seu Maké Seu Vêba Meu Mepe Nosso Jufé Seu Ziba Meu dabe Esse Jomu Seu Mefa Meu Chivú Seu Sogué Meu Nêque Esse Tadu Nosso Pilé Seu Puva Meu Jomu Meu Zobe Seu Maké Seu Bapa Meu Nêque



**Anexo 7- Tempo individual da 2<sup>a</sup> atividade experimental**

Grupo 1	GRAMATICAL	AGRAMATICAL
Bebê 1	5,683333333	7,483333333
Bebê 2	6,85	8,316666667
Bebê 3	3,3	8,3
Bebê 4	7,716666667	7,866666667
Bebê 5	6,266666667	7,8
Bebê 6	9,6	10,36666667
Bebê 7	13,28333333	14,28333333
Bebê 8	5,083333333	3,433333333
Bebê 9	2,616666667	3,7
Grupo 2		
Bebê 1	7,566666667	11,86666667
Bebê 2	4,716666667	4,166666667
Bebê 3	4,05	3,416666667
Bebê 4	2,933333333	2,816666667
Bebê 5	11,01666667	9,95
Bebê 6	12,81666667	13,06666667
Bebê 7	6,666666667	8,55
Bebê 8	6,6	9,916666667
Bebê 9	3,933333333	7,8
<b>MÉDIA</b>	<b>6,705555556</b>	<b>7,95</b>