

# Contribuições para o Atlas do Projeto AMPER – Norte: Variedade Linguística de Baião (PA).

Rosinele Lemos e Lemos (UFPA-SEDUC)

Programa de Pós –Graduação em Letras

UFPA

Belém, Brazil

e-mail: [rosinele\\_lemos@terra.com.br](mailto:rosinele_lemos@terra.com.br)

Regina Cruz (UFPA-CNPq)

Faculdade de Letras

UFPA/CNPq

Belém, Brazil

e-mail: [regina@ufpa.br](mailto:regina@ufpa.br)

**RESUMO** - O objetivo deste trabalho é caracterizar prosodicamente a variedade linguística de Baião (PA). Adota-se aqui a metodologia do Projeto AMPER. Os resultados preliminares indicam que as variações de F0 contribuem para distinguir enunciados declarativos e interrogativos da variedade investigada.

**Palavras-chave:** Projeto AMPER; Prosódia; Atlas Dialectal; Português Brasileiro.

**Abstract**— This work aims to characterize prosodically one Amazon Brazilian Portuguese linguistic variety spoken in Baião city. We adopted here the AMPER Project's methodology. The first results indicate that the F0 variations contribute to distinguish declarative and interrogative sentences.

**Keywords:** *Amper Project; Prosody; Dialectal Atlas; Brazilian Portuguese.*

## INTRODUÇÃO

O presente trabalho objetiva apresentar os resultados preliminares da pesquisa de Dissertação de Mestrado de Lemos (em andamento), vinculada ao projeto internacional AMPER<sup>1</sup> (*Atlas Multimédia Prosodique de l'Espace Romain*). Apresentam-se aqui particularmente os resultados obtidos com o tratamento dos dados de um informante do sexo feminino e de baixa escolaridade (BF91). A Universidade Federal do Pará (UFPA) participa desse projeto desde 2007, com a responsabilidade de confeccionar o Atlas Prosódico do Português do Norte do Brasil (Amper-Norte). Atualmente, estão formados *corpora* nas seguintes localidades paraenses: a) Cameté [10]; b) Belém [11], [12], [1]; c) Bragança [2]; d) Abaetetuba [8] Mosqueiro [6]; f) Currálinho [5]; g) Mocajuba [3]; e h) Baião [7] cujos dados compõem a referida pesquisa.

As análises dos dados de BF91 foram realizadas considerando os parâmetros físicos: frequência fundamental F0, intensidade e duração com o objetivo de verificar qual(is) parâmetro(s) contribui(em) na distinção entre as modalidades frasais declarativas e interrogativas na variedade investigada.

Este trabalho está dividido nas seguintes secções: secção I - Projeto AMPER-Norte; secção II - Metodologia do projeto

AMPER; secção III Variedade linguística investigada de Baião PA (Lemos, em andamento); secção IV - Procedimentos metodológicos; secção V - Resultados preliminares; secção VI - Conclusão; e secção VII - Referências bibliográficas.

## I. PROJETO AMPER – NORTE

O projeto AMPER-Norte está diretamente ligado ao projeto de pesquisa europeu AMPER, coordenado pelos professores Michel Contini e Jean-Pierre Lai, do Centro de Dialectologie da Universidade de Grenoble 3 (França); Antonio Romano da Universidade de Turim (Itália) e Albert Rillard do CNRS, Paris (França) e do qual fazem parte onze<sup>2</sup> outras instituições, além da UFPA, responsáveis por investigar o português (AMPER-POR).

O projeto AMPER tem como objetivo principal a organização prosódica das variedades faladas no espaço dialetal românico. O AMPER planeja também a disponibilização *online* do *corpus* com a intenção de possibilitar futuras investigações a diversos níveis da análise linguística.

A pesquisa relativa à variação prosódica do Português (AMPER-POR<sup>3</sup>) é coordenada pela Professora Lurdes de Castro Moutinho do Centro de Investigação de Línguas e Culturas da Universidade de Aveiro (Portugal).

O projeto **Amper-Norte** é sediado no laboratório de Ciência e Tecnologia da Fala do *Campus* Universitário de Cameté (CUNTINS) da UFPA e conta com a infraestrutura desta para a execução das atividades. O projeto em questão já produziu duas monografias de conclusão de Curso [11] e [2]; quatro Dissertações de Mestrado [10], [5], [6] e [8] e dois planos de Iniciação Científica [12] e [1]. A coordenadora do

<sup>1</sup> [http://w3.u-grenoble3.fr/dialecto/AMPER/AMPERfr\\_fichiers/frame.htm](http://w3.u-grenoble3.fr/dialecto/AMPER/AMPERfr_fichiers/frame.htm)

Identify applicable sponsor/s here. (*sponsors*).

<sup>2</sup> Universidade de Aveiro (Portugal), Universidade dos Açores (Portugal), Universidade da Madeira (Portugal), Universidade Federal do Amazonas (UFAM), Universidade Federal da Paraíba (UFPB), Universidade Federal do Rio de Janeiro (UFRJ), Universidade Federal Fluminense (UFF), Pontifícia Universidade Católica de São Paulo (PUCSP), Universidade Federal de Minas Gerais (UFMG), Universidade Federal do Paraná (UFPR), Universidade Federal do Espírito Santo (UFES) e a Universidade Federal de Santa Catarina (UFSC).

<sup>3</sup> [www2.ii.ua.pt/cidlc/gcl/AMPER-POR.htm](http://www2.ii.ua.pt/cidlc/gcl/AMPER-POR.htm) / [www2.ii.ua.pt/cidlc/gcl/](http://www2.ii.ua.pt/cidlc/gcl/)

projeto é bolsista de Produtividade do CNPq (PQ2) desde 2009<sup>4</sup>.

## II. METODOLOGIA DO PROJETO AMPER

Considerando que o projeto AMPER compreende uma análise contrastiva dos dialetos estudados, o *corpus* gravado era composto de seis repetições de 66 frases que foram estruturadas obedecendo às mesmas restrições fonéticas e sintáticas, a fim de se manter o mesmo padrão para todos os *corpora* coletados nas três variedades do português que fazem parte do projeto AMPER-POR. Atualmente, este *corpus* foi ampliado para 102 frases, totalizando 612 sinais sonoros.

Sintaticamente, as frases foram montadas de forma a apresentar Sujeito - Verbo – complemento (SVC) e suas expansões com a inclusão de Sintagmas Preposicionais. Todas as frases possuem apenas: 1) três personagens: Renato, pássaro e bisavô; 2) três sintagmas adjetivais: nadador, bêbado e pateta; 3) três sintagmas preposicionados indicadores de lugar: de Mônaco, de Veneza e de Salvador; e 4) um único verbo: gostar. Com relação à entoação, elas foram concebidas de modo a contemplar as modalidades declarativas e interrogativas. O *corpus* também foi formado considerando as três pautas acentuais do português: oxítone, paroxítone e proparoxítone.

Conforme determina o projeto geral - AMPER-POR - os informantes selecionados para a formação dos *corpora* devem atender aos seguintes critérios: 1) ser nativo ou ter vivido a maior parte de sua vida na localidade alvo; 2) ter idade acima de trinta anos; e 3) possuir escolaridade de ensino fundamental, médio ou superior. Trata-se, portanto, de uma amostra estratificada. Cada informante recebe um código, que contém informações sobre seu perfil. A partir desses critérios, são selecionados seis informantes: três homens e três mulheres. No momento da gravação do *corpus*, a cada informante são pedidas seis repetições da série de frases do *corpus* (em ordem aleatória), sendo selecionadas para análise acústica as três melhores repetições, a fim de se poderem estabelecer médias nos diversos parâmetros acústicos: duração, frequência fundamental e intensidade.

O material gravado sofre cinco etapas de tratamento: a) codificação das repetições; b) segmentação vocálica dos sinais selecionados no programa PRAAT 5.0; c) aplicação do *script* Praat; d) seleção das três melhores repetições e; e) aplicação da interface Matlab para se obter as médias dos parâmetros das três melhores repetições.

No caso da codificação das repetições, retoma-se o código do informante, contendo o seu perfil, acrescenta-se o código de cada frase já estabelecido pelo projeto AMPER, com as indicações sintáticas, fonéticas e prosódicas, por último acrescenta-se um número de ordem cronológica da repetição.

Para o trabalho de segmentação fonética, utiliza-se o programa PRAAT. Apenas um nível de segmentação fonética é criado, denominado de <vogais>. O *script* PRAAT lê como códigos apenas a letra “v” (indica vogais plenas) e a letra “f”

(indica vogais fracas ou elididas). Durante a segmentação fonética são estabelecidas as escalas de *pitch* adequadas para a análise de cada informante.

Concluída a segmentação fonética de todos os sinais de áudio - 396 do *corpus* antigo e 612 do *corpus* expandido - de cada informante, passa-se à aplicação do *script praata*. O *script praata* é aplicado a cada uma das 396 (do *corpus* antigo) ou 612 (do *corpus* expandido) repetições obtidas com a codificação da gravação original. A aplicação desse *script* gera um arquivo.TXT contendo as medidas dos parâmetros acústicos (intensidade, frequência fundamental, duração e formantes) das vogais de cada repetição.

Previamente à análise acústica na interface Matlab, selecionam-se as três melhores repetições de cada frase em termos de qualidade sonora e de similaridade de distribuição de vogais plenas (v) e elididas (f).

A aplicação da interface Matlab fornece a média dos parâmetros físicos – F0, duração e intensidade – em um arquivo fono.txt das três repetições de cada frase e das duas modalidades. A interface gera mais outros arquivos em formato de imagem contendo gráficos das médias de F0, duração e intensidade de cada modalidade individualmente, assim como gráficos comparativos de ambas as modalidades. A interface gera igualmente arquivos ton contendo uma síntese de cada modalidade sem a parte segmental.

## III. VARIEDADE LINGÜÍSTICA INVESTIGADA DE BAIÃO PA (LEMONS EM ANDAMENTO)

Os dados deste trabalho referem-se à variedade linguística do português falado no município de Baião (PA) e são considerados inéditos.

O *corpus* está de acordo com a metodologia do projeto AMPER quanto à organização e tratamento dos dados. Em Baião foram gravadas as 66 frases do *corpus* antigo<sup>5</sup>. Os informantes são nativos da localidade, com idade acima de trinta anos e níveis de escolaridade fundamental, médio e superior. A gravação ocorreu na própria casa dos informantes para garantir a espontaneidade do discurso. Utilizou-se gravador digital PMD660 Marant microfone Shure dinâmico e de cabeça para a captura do áudio, e computador notebook para a projeção das imagens e controle da qualidade das gravações no software Soundforge.

As figuras, em slides, formadoras das 66 frases foram exibidas pelo computador no programa PowerPoint aos informantes, sem apresentar-lhes suas formas escritas. Cada informante repetiu seis vezes a série de frases (em ordem aleatória), no momento da gravação sem interrupção até a última frase. Ao todo foram obtidos seis sinais sonoros de 6h33min15s de gravação. A taxa de amostragem de cada sinal é de 44.100 Hz, 16 bits, sinal mono. Cada informante recebeu um código com informações sobre seu perfil.

4

Processos 303455/2008-0 e 308092/2011-2.

5 O *corpus* ampliado de 102 frases está sendo utilizado por [8], [5] e [3] (em andamento).

TABELA I. - CÓDIGO, BAIRRO E DURAÇÃO DO CORPUS GRAVADO DOS INFORMANTES DE BAIÃO (PA).

Informante	Bairro	Duração de Gravação
BF91	Limão	1 h, 23 min e 39 seg.
BF92	Cumbucão	1 h,43min e 36 seg.
BF93	Maracanã	48 min e 58 seg.
BF94	Centro	53 min e 50 seg
BF95	Centro	39 min e 29 seg
BF96	Limão	33 min e 36 seg

Até o presente momento foram explorados apenas os dados do informante BF91, como veremos no item seguinte.

#### IV. PROCEDIMENTOS METODOLÓGICOS

O material gravado do informante BF91 sofreu seis etapas de tratamento: a) codificação das repetições; b) isolamento das repetições em arquivos de áudio individuais; c) segmentação fonética dos sinais selecionados no programa PRAAT 5.0; d) aplicação do *script* praat; e) seleção das três melhores repetições e; f) aplicação da interface Matlab para se obter as médias dos parâmetros das três melhores repetições.

Na codificação das repetições desse informante acrescentou-se o código de cada frase com as indicações sintáticas, fonéticas e prosódicas, e um número de ordem cronológica da repetição para em seguida isolar as 396 frases do sinal original em um arquivo sonoro específico.

Na segmentação fonética utilizou-se o programa PRAAT e estabeleceu-se a escala de *pitch* que ficou entre 120 Hz a 350 Hz.

O *script praat* foi aplicado a cada uma das 396 repetições que gerou um arquivo.TXT contendo as medidas dos parâmetros acústicos (frequência fundamental, duração e intensidade) das vogais de cada repetição.

Antes de se proceder a análise acústica na interface Matlab, foram selecionadas as três melhores repetições de cada frase em termos de qualidade sonora e de similaridade de distribuição de vogais plenas (v) e elididas (f).

A aplicação da interface Matlab forneceu a média dos parâmetros físicos – F0, duração e intensidade – em um arquivo fono.txt das três repetições de cada frase e das duas modalidades. A interface gerou mais outros arquivos em formato de imagem contendo gráficos das médias de F0, duração e intensidade de cada modalidade individualmente, assim como gráficos comparativos de ambas as modalidades. A interface gerou igualmente arquivos ton contendo uma síntese de cada modalidade sem a parte segmental.

Para este estudo, considerou-se apenas os dados fornecidos para as frases com sintagmas nominais finais simples contendo 10 vogais, a saber: “O **pássaro** gosta do **pássaro**” (pwp), “O **Renato** gosta do **Renato**” (twk) e “O **bisavô** gosta do **bisavô**” (kwk), porque representam as três pautas acentuais do português (proparoxítona, paroxítona e oxítona).

#### V. RESULTADOS PRELIMINARES

A análise preliminar incidiu sobre as variações dos parâmetros físicos controlados – F0, duração e intensidade – sobre as sílabas tônicas do núcleo dos sintagmas nominais para verificar a hipótese, base do projeto Norte Vogais, de que as variações mais importantes de F0 ocorrem justamente na sílaba tônica do núcleo do sintagma final do enunciado analisado, como constatados para as variedades de Belém [4] e Cameté [10].

Os resultados de Belém [4] e de Cameté [10] demonstraram que a duração atua como parâmetro complementar na distinção das modalidades declarativas e interrogativas das sentenças analisadas, mas a intensidade não se apresenta como um parâmetro importante na distinção das modalidades em questão.

Os resultados de BF91 indicam também que as variações de F0 mais importantes ocorrem justamente na sílaba tônica do sintagma nominal final do enunciado (Figura 1).

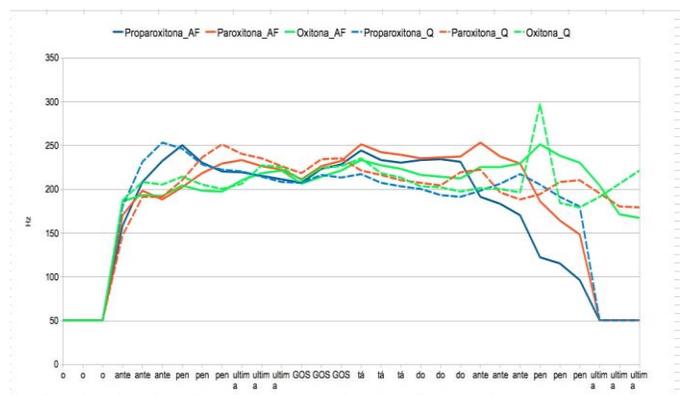


Fig.1 - Comparação entre a média de variação de F0 nas sentenças pwp (azul), twt (laranja) e kwk (verde) em ambas as modalidades – declarativa (linha plena) e interrogativa (linha pontilhada), faladas por um locutor do sexo feminino com baixa escolaridade de Baião.

As médias de variação de F0 nas sentenças pwp “O **pássaro** gosta do **pássaro**” (azul), twt “O **Renato** gosta do **Renato**” (laranja) e kwk “O **bisavô** gosta do **bisavô**” (verde), representam as três pautas acentuais do português: proparoxítona, paroxítona e oxítona, respectivamente. Observaram-se os contornos de F0 de cada uma dessas frases nas duas modalidades alvos: declarativa (em linha plena) e interrogativa (linha pontilhada). A posição do sujeito encontra-se no início dos vocábulos.

Segundo as análises, as variações das médias do movimento da curva melódica de F0 mais importantes concentram-se na antepenúltima sílaba (“O **pássaro** gosta do **pássaro**”), na penúltima (“O **Renato** gosta do **Renato**”) e na última (“O **bisavô** gosta do **bisavô**”). Nota-se igualmente que o contorno realizado apresenta-se em forma de “pinça”, padrão esperado nas variedades do português do Norte do Brasil. Desse modo, a hipótese apresentada por [9] é corroborada com esses resultados, pois é na última sílaba tônica da sentença que se evidencia o movimento mais importante de F0.

Apesar da variedade de Baião reforçar a hipótese defendida por [4] e [10] de que F0 é um parâmetro físico fundamental na distinção das modalidades em análise, os valores de duração não apresentam comportamentos semelhantes às variações de Belém e Cametá. (Figura 2).

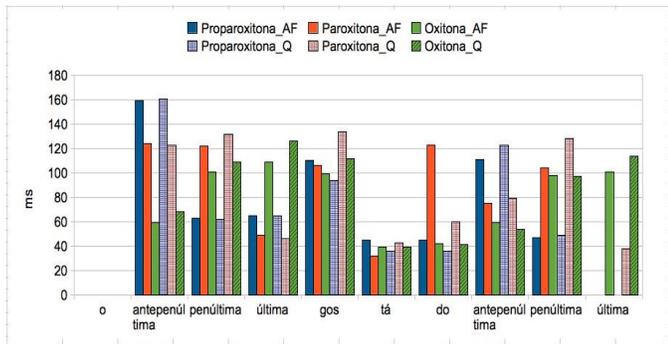


Fig. 2 - Comparação entre a média de variação de ms nas sentenças pwp (azul), twt (laranja) e kwk (verde) em ambas as modalidades, declarativa (coluna plena) e interrogativa (coluna com ranhuras), faladas por um locutor do sexo feminino com baixa escolaridade de Baião.

A média de variação de ms nas sentenças (Figura 2) O pássaro gosta do pássaro (pwp - azul); O Renato gosta do Renato (tw - laranja); e O bisavô gosta do bisavô (kwk - verde) para as três pautas acentuais do português: proparoxítona, paroxítona e oxítona, nas modalidades – declarativa (coluna plena) e interrogativa (coluna com ranhuras), cujo sujeito encontra-se no início do sintagma demonstra que as medidas de duração não são significativas para distinguir as duas modalidades frasais como observaram Cruz e Brito (2011) e Santo (2011). Da mesma forma, a intensidade também não se configura como um parâmetro relevante na distinção das duas modalidades na variedade de Baião (Figura 3).

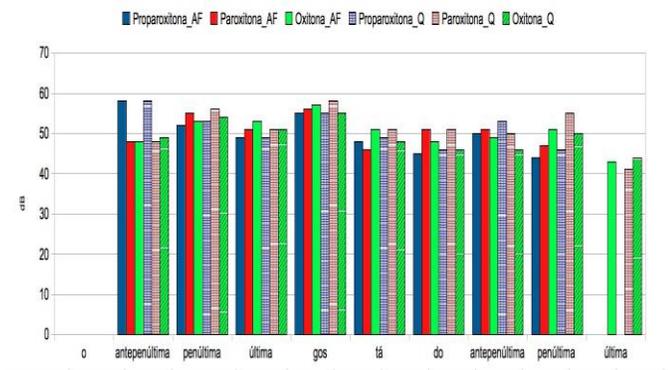


Fig. 3 - Comparação entre a média de variação de dB nas sentenças pwp (azul), twt (laranja) e kwk (verde) em ambas as modalidades, declarativa (coluna plena) e interrogativa (coluna com ranhuras), faladas por um locutor do sexo feminino com baixa escolaridade de Baião.

A média de variação de dB nas sentenças O pássaro gosta do pássaro – (azul), – O Renato gosta do Renato – (laranja) e - O bisavô gosta do bisavô (verde) – que representam as três pautas acentuais do português: proparoxítona, paroxítona e

oxítona, em ambas as modalidades – declarativa (coluna plena) e interrogativa (coluna com ranhuras), com sujeito no início do enunciado não apresenta diferença significativa para distinguir as duas modalidades frasais pelo parâmetro intensidade. Os resultados de Baião acompanham aqueles obtidos em Belém por [4] e em Cametá por [10] de que a intensidade não é um parâmetro relevante para distinguir enunciados declarativos e interrogativos.

## VI. CONCLUSÃO

Os resultados preliminares da Dissertação de Mestrado de LEMOS (em andamento) do informante BF91 quanto aos parâmetros físicos de intensidade, duração e frequência fundamental (F0), relacionados à pauta acentual do português para a variedade investigada, permitem concluir que apenas o parâmetro físico de frequência fundamental F0 é relevante para a distinção entre as modalidades declarativas e interrogativas do português falado no município. Os parâmetros de duração de ms e de intensidade dB não se configuram relevantes para a distinção das duas modalidades frasais na variedade linguística de Baião (PA).

## VII. REFERÊNCIAS

- [1] C. Brito. *Atlas prosódico multimídia do Português do Norte do Brasil – AMPER-POR: variedade linguística da zona rural de Belém (PA)*, Belém: UFPA/ILC/FALE, 2012 (Iniciação Científica).
- [2] F. C. Castilho. *Formação de Corpora para o Atlas Dialectal Prosódico Multimídia do Norte do Brasil: Variedade Linguística de Bragança (PA)*. Bragança: UFPA/Campus de Bragança/Faculdade de Letras, 2009. (Trabalho de Conclusão de Curso em Letras).
- [3] M. S. Costa. *Atlas Prosódico Multimídia do Município de Mocajuba (PA)*, Belém: UFPA/ILC/CML, em andamento (Dissertação de Mestrado).
- [4] R. Cruz e C. Brito. *Atlas Prosódico Multimídia da Cidade de Belém (PA): uma visão geral*. Comunicação oral apresentada durante o V Congresso de Fonética Experimental. Cáceres (Espanha): Universidad de Extremadura, 25 a 28 de outubro de 2011.
- [5] J. Freitas. *Atlas Prosódico Multimídia do Município da ilha do Marajó (PA)*, Belém: UFPA/ILC/CML, 2013. (Dissertação de Mestrado).
- [6] E. Guimarães. *Atlas Prosódico Multimídia da ilha de Mosqueiro (PA)*, Belém: UFPA/ILC/CML, 2013. (Dissertação de Mestrado).
- [7] R. Lemos. *Atlas Prosódico Multimídia do Município de Baião (PA)*, Belém: UFPA/ILC, CML, em andamento (Dissertação de Mestrado).
- [8] I. Remédios, *Atlas Prosódico Multimídia do Município de Abaetetuba (PA)*, Belém: UFPA/ILC/CML, 2013. (Dissertação de Mestrado).
- [9] I. Santo e R. Cruz. *Atlas Prosódico Multimídia do Município de Cametá (PA): uma visão geral*. Comunicação oral apresentada durante o III Colóquio De Prosódia da Fala. Belo Horizonte (MG – Brasil): Universidade Federal de Minas Gerais, 6 a 8 de junho de 2011.
- [10] I. Santo. *Atlas Prosódico Multimídia do Município de Cametá (PA)*, Belém: UFPA/ILC/CML, 2011. (Dissertação de Mestrado).
- [11] JR. M F. Santos. *Formação de corpora para o Atlas Dialectal Prosódico Multimídia do Norte do Brasil: variedade linguística de Belém*. Belém: UFPA/ILC/FALE, 2008. (Trabalho de Conclusão de Curso em Letras).
- [12] A. Silva. *Atlas prosódico multimídia do Português do Norte do Brasil – AMPER-POR: variedade linguística da zona rural de Belém (PA)*. Belém: UFPA. 2011. (Plano PIBIC/CNPq).