UNIVERSIDADE ESTADUAL PAULISTA – UNESP
Faculdade de Ciências e Letras - Câmpus de Araraquara

JUAN PRETE TOJEIRA RAMOS

PROSÓDIA COMPUTACIONAL DO PORTUGUÊS BRASILEIRO:

a entoação declarativa neutra gerada por um sistema de síntese de fala baseado em Inteligência Artificial (IA)

Araraquara – SP
2025

JUAN PRETE TOJEIRA RAMOS

PROSÓDIA COMPUTACIONAL DO PORTUGUÊS BRASILEIRO:

a entoação declarativa neutra gerada por um sistema de síntese de fala baseado em Inteligência Artificial (IA)

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), por meio da Faculdade de Ciências e Letras (FCL), Câmpus de Araraquara, para a obtenção do título de Mestre em Linguística e Língua Portuguesa.

Área de Concentração: Linguística e Língua Portuguesa.

Linha de Pesquisa: Teoria, Descrição e Análise de Línguas Naturais.

Orientadora: Profa. Dra. Gladis Massini-Cagliari.

Agência financiadora: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Araraquara – SP

2025

T646p

Tojeira-Ramos, Juan Prete
Prosódia computacional do português brasileiro : a entoação declarativa neutra gerada por um sistema de síntese de fala baseado em Inteligência Artificial (IA) / Juan Prete Tojeira-Ramos. -- Araraquara, 2025
148 p. : il.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências e Letras, Araraquara
Orientadora: Gladis Massini-Cagliari

1. Fonética. 2. Fonologia. 3. Inteligência artificial. 4. Geração de linguagem natural (Computação). 5. Entonação (Fonética). I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

IMPACTO POTENCIAL DESTA PESQUISA

A pesquisa contribui para a produção científica no campo da Fonética, da Fonologia e da Linguística Computacional e estabelece um diálogo com o Processamento de Linguagem Natural (PLN). Ao demonstrar uma correspondência sistemática entre os eventos tonais observados na fala sintética e os padrões fonológicos do português brasileiro na modalidade declarativa neutra, o trabalho amplia o conhecimento sobre a interface entre os processos estatísticos de modelos de síntese de fala e as regularidades linguísticas da entoação dessa variedade. Tal achado oferece fundamentos teóricos para a formulação de hipóteses acerca da representação prosódica em ambientes computacionais e para o refinamento de modelos analíticos que integrem as pistas fonéticas e as categorias fonológicas.

No plano técnico e inovador, o estudo fornece critérios metodológicos e métricas aplicáveis à avaliação da qualidade entoacional em produtos de síntese de fala, bem como orientações para o planejamento de controles prosódicos mais precisos em mecanismos derivados de algoritmos matemáticos e de redes neurais artificiais. A descrição detalhada de procedimentos de extração e de verificação da frequência fundamental (F_0), combinada com uma notação melódica compatível com uma interpretação fonológica, possibilita aos engenheiros e cientistas da computação a otimização de módulos prosódicos, o que resulta em vozes sintéticas com mais naturalidade e coerência entoacional.

No âmbito social e inclusivo, os resultados apoiam a implementação de iniciativas destinadas à democratização da comunicação. A validação de contornos entoacionais consistentes na fala sintética, para a declaração neutra, indica o potencial do recurso para a criação de leitores automáticos, de tecnologias assistivas e de interfaces de voz que propiciem um adequado entendimento por parte de pessoas com deficiência visual ou dificuldade leitora. Nos cenários de saúde e de reabilitação, tais aplicações recebem um respaldo técnico, ao passo que, nos materiais didáticos formados por arquivos de áudio, a entoação declarativa neutra pode ser aprimorada, o que tende a beneficiar o processo de ensino e de aprendizagem e, em consequência, a experiência dos estudantes em relação ao conteúdo programático.

No domínio econômico, a investigação revela a existência de vantagens estratégicas para os fornecedores de serviços de síntese de fala que adotam os parâmetros prosódicos validados por procedimentos empíricos. O emprego de controles entoacionais robustos pode reduzir os custos provenientes de sucessivas atualizações durante o desenvolvimento de produtos, melhorar a aceitação do público e promover as possibilidades de comercialização em assistentes virtuais e canais de atendimento ao consumidor. Ademais, o trabalho orienta a

elaboração de parâmetros de avaliação prosódica de produtos comerciais e favorece a constituição de diretrizes técnicas e de certificações industriais.

Quanto à internacionalização e à inserção local, regional e nacional, o estudo proporciona à comunidade acadêmica uma interlocução direta com diversas pesquisas brasileiras e internacionais sobre as temáticas relacionadas ao trabalho, ao mesmo tempo em que valoriza as especificidades da variedade do português estudada. A metodologia proposta permite a replicação da investigação em outros contextos linguísticos e fomenta a cooperação entre as universidades, os laboratórios de pesquisa e as empresas de tecnologia digital, de modo a estimular novas parcerias científicas no Brasil e no exterior.

Em termos educacionais e culturais, a pesquisa expande os repertórios analíticos e disponibiliza um conjunto de subsídios para a docência em cursos de graduação e programas de pós-graduação em Linguística e Ciência da Computação. A documentação do padrão entoacional de enunciados declarativos neutros da variedade brasileira do português auxilia na preservação e na divulgação de conhecimentos sobre as características prosódicas nacionais, além de assegurar o debate entre a pesquisa teórica e a aplicação da fala sintética em plataformas midiáticas e escolares.

Por último, observa-se uma convergência com os Objetivos de Desenvolvimento Sustentável (ODS), definidos pela Organização das Nações Unidas (ONU) para a Agenda 2030, sobretudo no que diz respeito aos eixos de ação 3 (“Saúde e Bem-Estar”), 4 (“Educação de Qualidade”), 9 (“Indústria, Inovação e Infraestrutura”), 10 (“Redução das Desigualdades”) e 17 (“Parcerias e Meios de Implementação”). Dentre os aspectos que evidenciam a referida convergência, destacam-se a valorização da inclusão social e da acessibilidade comunicativa, a eficiência dos serviços de saúde comunitária e a construção de parcerias direcionadas à geração de um impacto social em larga escala.

POTENTIAL IMPACT OF THIS RESEARCH

The research contributes to scientific production in the fields of Phonetics, Phonology, and Computational Linguistics and establishes a dialogue with Natural Language Processing (NLP). By demonstrating a systematic correspondence between the tonal events observed in synthetic speech and the phonological patterns of Brazilian Portuguese in the neutral declarative mode, the work expands knowledge about the interface between the statistical processes of speech synthesis models and the linguistic regularities of intonation in this variety. This finding provides theoretical foundations for formulating hypotheses about prosodic representation in computational environments and for refining analytical models that integrate phonetic cues and phonological categories.

On a technical and innovative level, the study provides methodological criteria and metrics applicable to the evaluation of intonational quality in speech synthesis products, as well as guidelines for planning more accurate prosodic controls in mechanisms derived from mathematical algorithms and artificial neural networks. The detailed description of procedures for extracting and verifying fundamental frequency (F0), combined with a melodic notation compatible with phonological interpretation, enables engineers and computer scientists to optimize prosodic modules, resulting in synthetic voices with greater naturalness and intonational coherence.

In the social and inclusive sphere, the results support the implementation of initiatives aimed at democratizing communication. The validation of consistent intonational contours in synthetic speech for neutral statements indicates the potential of this resource for the creation of automatic readers, assistive technologies, and voice interfaces that enable adequate understanding by people with visual impairments or reading difficulties. In health and rehabilitation scenarios, such applications receive technical support, while in teaching materials consisting of audio files, neutral declarative intonation can be improved, which tends to benefit the teaching and learning process and, consequently, the students' experience in relation to the program content.

In the economic domain, research reveals strategic advantages for speech synthesis service providers who adopt prosodic parameters validated by empirical procedures. The use of robust intonational controls can reduce the costs of successive updates during product development, improve public acceptance, and promote commercialization opportunities in virtual assistants and customer service channels. In addition, the work guides the development of prosodic evaluation parameters for commercial products and favors the establishment of

technical guidelines and industrial certifications.

In terms of internationalization and local, regional, and national integration, the study provides the academic community with direct access to various Brazilian and international research projects on work-related topics, while also highlighting the specific characteristics of the variety of Portuguese studied. The proposed methodology allows the replication of research in other linguistic contexts and fosters cooperation between universities, research laboratories, and digital technology companies, in order to stimulate new scientific partnerships in Brazil and abroad.

In educational and cultural terms, the research expands analytical repertoires and provides a set of resources for teaching undergraduate and graduate courses in Linguistics and Computer Science. The documentation of the intonational pattern of neutral declarative utterances in the Brazilian variety of Portuguese helps to preserve and disseminate knowledge about national prosodic characteristics, in addition to ensuring debate between theoretical research and the application of synthetic speech in media and school platforms.

Finally, there is convergence with the Sustainable Development Goals (SDGs) defined by the United Nations (UN) for the 2030 Agenda, especially with regard to action areas 3 (“Good Health and Well-Being”), 4 (“Quality Education”), 9 (“Industry, Innovation and Infrastructure”), 10 (“Reduced Inequalities”), and 17 (“Partnerships for the Goals”). Among the aspects that highlight this convergence are the promotion of social inclusion and communicative accessibility, the efficiency of community health services, and the establishment of partnerships aimed at generating large-scale social impact.

JUAN PRETE TOJEIRA RAMOS

PROSÓDIA COMPUTACIONAL DO PORTUGUÊS BRASILEIRO:

a entoação declarativa neutra gerada por um sistema de síntese de fala baseado em Inteligência Artificial (IA)

Dissertação apresentada à Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), por meio da Faculdade de Ciências e Letras (FCL), Câmpus de Araraquara, para a obtenção do título de Mestre em Linguística e Língua Portuguesa, com o financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Linha de Pesquisa: Teoria, Descrição e Análise de Línguas Naturais.

Data da defesa: 10/12/2025.

Banca Examinadora:

_____

Presidente e Orientadora: Profa. Dra. Gladis Massini-Cagliari
UNESP - Faculdade de Ciências e Letras - Câmpus de Araraquara

_____

Membro Titular: Profa. Dra. Larissa Cristina Berti
UNESP - Faculdade de Filosofia e Ciências - Câmpus de Marília

_____

Membro Titular: Profa. Dra. Adelaide Hercília Pescatori Silva
UFPR - Setor de Ciências Humanas - Câmpus de Curitiba

_____

Membro Suplente: Profa. Dra. Maíra Sueco Maegava Córdula
UFU - Instituto de Letras e Linguística - Câmpus de Santa Mônica

_____

Membro Suplente: Profa. Dra. Débora Aparecida dos Reis Justo Barreto
Pesquisadora independente

Dedico este trabalho à minha mãe, Sirlene Aparecida Prete, cujo amor, afeto e dedicação são imprescindíveis para o êxito de minha carreira acadêmica, científica e profissional.

AGRADECIMENTOS

Gostaria de expressar meus agradecimentos especiais à Profa. Dra. Gladis Massini-Cagliari, responsável pela orientação durante o Mestrado Acadêmico, pelo acompanhamento cuidadoso e pelo incentivo contínuo ao longo do desenvolvimento da investigação. Também registro minha gratidão ao Programa de Pós-Graduação em Linguística e Língua Portuguesa da Faculdade de Ciências e Letras (FCL), da Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Câmpus de Araraquara, bem como ao Grupo de Pesquisa "Fonologia do Português: Arcaico & Brasileiro", coordenado por minha orientadora e pelo Prof. Dr. Luiz Carlos Cagliari, pelo suporte institucional e pelo estímulo à pesquisa.

Agradeço aos Profs. Drs. Gladis Massini-Cagliari, Juliana Simões Fonte, Débora Aparecida dos Reis Justo Barreto, Regiani Aparecida Santos Zacarias e Guilherme Duarte Garcia o ensino qualificado, a discussão acadêmica e o estímulo à reflexão crítica nas disciplinas realizadas no decorrer do curso. Ademais, reconheço os Profs. Drs. Ubiratã Kickhöfel Alves, Guilherme Duarte Garcia e Flaviane Romani Fernandes Svartman pela oportunidade de diálogo em eventos científicos brasileiros.

É imprescindível manifestar minha gratidão às Profas. Dras. Adelaide Hercília Pescatori Silva e Larissa Cristina Berti pela avaliação nas etapas de qualificação e de defesa, assim como às Profas. Dras. Maíra Sueco Maegava Córdula e Débora Aparecida dos Reis Justo Barreto pela participação como suplentes da banca de defesa.

Ainda dirijo meu reconhecimento ao corpo docente do Curso de Licenciatura em Letras do Instituto de Biociências, Letras e Ciências Exatas (Ibilce), Câmpus de São José do Rio Preto, com destaque aos primeiros orientadores, Profs. Drs. Douglas Altamiro Consolo e Erotilde Goreti Pezatti, pelo acompanhamento em meus estudos científicos iniciais. Nesse período, também registro o convívio com o Grupo de Pesquisa em Gramática Funcional (GPGF), coordenado pelos Profs. Drs. Erotilde Goreti Pezatti e Roberto Gomes Camacho.

Por fim, sou grato à minha família, em especial aos meus pais, Sirlene Aparecida Prete e Ricardo Tojeira Ramos, pelo companheirismo e pela confiança demonstrados desde o começo de minha carreira acadêmica, científica e profissional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

"[...] os elementos prosódicos não são simples enfeites fonéticos da linguagem oral, mas uma das maneiras que a linguagem tem de carregar significados" (Cagliari, 2002a, p. 46).

RESUMO

Esta dissertação investiga a entoação declarativa neutra no português brasileiro, gerada por um recurso de conversão de texto escrito em fala audível baseado em Inteligência Artificial (IA), desenvolvido pela empresa multinacional americana Google. Os objetivos gerais consistem em caracterizar as propriedades entoacionais de enunciados declarativos neutros sintéticos e compará-las com a entoação natural do português brasileiro. Especificamente, objetiva-se identificar os eventos tonais associados ao contorno de entoação dos enunciados declarativos neutros, determinar o padrão entoacional do conjunto de enunciados sintéticos e comparar os aspectos entoacionais desses enunciados com os resultados do estudo de Tenani (2002) sobre a entoação declarativa neutra do português brasileiro em um contexto de fala natural. A pesquisa, auxiliada pela Fonética Acústica, adota uma visão integrada entre a entoação e os domínios prosódicos, com base na Fonologia Entoacional (Ladd, 1996, 2008) e na Fonologia Prosódica (Nespor; Vogel, 1986, 2007), além de considerar os trabalhos que tratam da variação entoacional do português brasileiro. A análise se concentra em enunciados declarativos neutros produzidos por vozes sintéticas do produto investigado, com foco no parâmetro acústico da frequência fundamental (F_0). Para a identificação dos eventos tonais associados ao contorno de entoação dos enunciados declarativos neutros, conforme as diretrizes do sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015), são utilizadas a inspeção auditiva e a versão 6.4.04 do *software* Praat (Boersma; Weenink, 2024). Determina-se o padrão entoacional do conjunto de enunciados declarativos neutros com o auxílio da descrição dos tons atribuídos ao contorno melódico, e a análise entoacional dos dados de fala sintética é comparada com o estudo de Tenani (2002) e outras pesquisas sobre a entoação do português brasileiro. Hipotetiza-se que a entoação desses enunciados decorra da mesma organização fonológica da fala natural do português brasileiro, pois as arquiteturas de redes neurais artificiais são utilizadas para a identificação de padrões prosódicos recorrentes em amostras humanas e para a reprodução robusta de traços melódicos, ainda que por meio de mecanismos estatísticos. Os resultados indicam que os eventos tonais observados correspondem à gramática prosódica do português brasileiro, reforçam a alta densidade tonal dessa variedade linguística e destacam a importância da palavra fonológica na atribuição de acentos tonais. A pesquisa identifica um padrão entoacional predominante, com uma ascendência tonal no começo e uma descendência melódica no término dos enunciados. Ademais, o trabalho mostra que a fala sintética é semelhante à natural, no que se refere à declaração neutra, em termos de eventos tonais, e demonstra a capacidade do Google Cloud Text-to-Speech de gerar, com precisão, a estrutura entoacional dessa modalidade enunciativa no português brasileiro. Em suma, o estudo destaca a compatibilidade da entoação declarativa neutra sintética com a gramática prosódica de falantes proficientes do português brasileiro e contribui para o avanço da Linguística Computacional. Conclui-se que a fala sintética analisada dispõe de propriedades entoacionais similares às da fala natural na declaração neutra, o que evidencia um progresso significativo das tecnologias de síntese de fala baseadas em IA. Essa correspondência comprova que o sistema computacional investigado modela e reproduz a entoação declarativa neutra de maneira consistente e corrobora as hipóteses da pesquisa.

Palavras-chave: prosódia computacional; síntese de fala; entoação; português brasileiro.

ABSTRACT

This dissertation investigates neutral declarative intonation in Brazilian Portuguese, generated by a text-to-speech conversion tool based on Artificial Intelligence (AI), developed by the American multinational company Google. The general objectives are to characterize the intonational properties of synthetic neutral declarative utterances and compare them with the natural intonation of Brazilian Portuguese. Specifically, the objective is to identify the tonal events associated with the intonation contour of neutral declarative utterances, determine the intonational pattern of the set of synthetic utterances, and compare the intonational aspects of these utterances with the results of Tenani's (2002) study on the neutral declarative intonation of Brazilian Portuguese in a natural speech context. The research, aided by Acoustic Phonetics, adopts an integrated view of intonation and prosodic domains, based on Intonational Phonology (Ladd, 1996, 2008) and Prosodic Phonology (Nespor; Vogel, 1986, 2007), in addition to considering works that deal with intonational variation in Brazilian Portuguese. The analysis focuses on neutral declarative statements produced by synthetic voices of the investigated product, focusing on the acoustic parameter of fundamental frequency (F_0). To identify the tonal events associated with the intonation contour of neutral declarative statements, according to the guidelines of the P-ToBI system (Frota; Oliveira, P.; Cruz; Vigário, 2015), auditory inspection and version 6.4.04 of the Praat software (Boersma; Weenink, 2024) are used. The intonation pattern of the set of neutral declarative utterances is determined with the aid of the description of the tones assigned to the melodic contour, and the intonation analysis of the synthetic speech data is compared with the study by Tenani (2002) and other research on Brazilian Portuguese intonation. It is hypothesized that the intonation of these utterances stems from the same phonological organization present in natural Brazilian Portuguese speech, since artificial neural network architectures are used to identify recurring prosodic patterns in human samples and to robustly reproduce melodic features, albeit through statistical mechanisms. The results indicate that the tonal events observed correspond to the prosodic grammar of Brazilian Portuguese, reinforce the high tonal density of this linguistic variety, and highlight the importance of the phonological word in the assignment of pitch accents. The research identifies a predominant intonational pattern, with a tonal rising at the beginning and a melodic falling at the end of utterances. Furthermore, the study shows that synthetic speech is similar to natural speech in terms of neutral utterances, in terms of tonal events, and demonstrates the ability of Google Cloud Text-to-Speech to accurately generate the intonational structure of this type of utterance in Brazilian Portuguese. In short, the study highlights the compatibility of synthetic neutral declarative intonation with the prosodic grammar of proficient speakers of Brazilian Portuguese and contributes to the advancement of Computational Linguistics. It is concluded that the synthetic speech analyzed has intonational properties similar to those of natural speech in the neutral declaration, which shows significant progress in AI-based speech synthesis technologies. This correspondence proves that the investigated computer system models and reproduces neutral declarative intonation consistently and corroborates the research hypotheses.

Keywords: computational prosody; speech synthesis; intonation; Brazilian Portuguese.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS[1]

Gerais

ONU	Organização das Nações Unidas
ODS	Objetivos de Desenvolvimento Sustentável
pt-BR	Português brasileiro
et al.	e outros
/ ... /	Transcrição fonológica
| |	Delimitação visual de palavras e de sílabas ortográficas

Ciência da Computação

IA	Inteligência Artificial
PLN	Processamento de Linguagem Natural
TTS	*Text-to-Speech* (conversão de texto em fala)
API	*Application Programming Interface* (Interface de Programação de Aplicação)

Fonética Acústica

dB	Decibéis
ms	Milissegundos
s	Segundos
F_0	Frequência fundamental
Hz	Hertz
<	Pico atrasado (*delayed peak*)
>	Pico adiantado (*early peak*)
¡	Pico elevado (*raised peak*)

[1] Nesta lista, são elencadas as principais abreviaturas, siglas e símbolos utilizados no trabalho.

Fonologia Entoacional Autossegmental e Métrica

P-ToBI	*Portuguese Tones and Break Indices*
H	Tom alto (*high tone*)
L	Tom baixo (*low tone*)
+	Combinação de dois tons no mesmo evento tonal
*	Alvo tonal associado a uma proeminência melódica
H*	Acento tonal alto (*high pitch accent*)
L*	Acento tonal baixo (*low pitch accent*)
H+L*	Acento tonal descendente (*falling pitch accent*)
L*+H/L+H*	Acento tonal ascendente (*rising pitch accent*)
H-	Acento frasal alto (*high phrasal accent*)
L-	Acento frasal baixo (*low phrasal accent*)
%	Símbolo de fronteira entoacional (*boundary tone*)
L%	Tom de fronteira baixo (*low boundary tone*)

Fonologia Prosódica

U	Enunciado fonológico (*phonological utterance*)
I	Frase entoacional (*intonational phrase*)
ϕ	Frase fonológica (*phonological phrase*)
C	Grupo clítico (*clitic group*)
PWG	Grupo de palavra prosódica (*prosodic word group*)
ω	Palavra fonológica (*phonological word*)
Σ	Pé métrico (*foot*)
σ	Sílaba (*syllable*)

SUMÁRIO

1 INTRODUÇÃO

As últimas décadas têm sido marcadas por diversas inovações eletrônicas e computacionais. No contexto contemporâneo, observa-se a implementação cada vez mais frequente de sistemas de Inteligência Artificial (IA), capazes de permitir que uma máquina seja integrada "a uma sociedade para executar tarefas competitivas que exigem processos cognitivos e se comunicar com outras entidades da sociedade por meio da troca de mensagens com alto conteúdo de informações e representações mais curtas" (Abbass, 2021, p. 95, tradução nossa).[2] Nesse sentido, a IA revoluciona a pesquisa científica com a automatização de tarefas, a análise de vastos volumes de dados e a aceleração de descobertas, fatores que contribuem para a ampliação do conhecimento e fortalecem a interação entre os seres humanos e as máquinas (Limongi, 2024), além da identificação de padrões estatísticos e da execução de operações complexas (Escovedo; Koshiyama, 2020; Kalinowski *et al.*, 2023). À medida que as investigações e as iniciativas em tecnologia de fala se aprimoram, uma parcela desses sistemas é programada para simular a voz humana e dialogar com os consumidores de diferentes países e culturas.

De modo geral, a síntese de fala é compreendida como a "produção de fala por máquinas, por meio da fonetização automática das frases a serem pronunciadas" (Dutoit, 1997, p. 13, grifos no original, tradução nossa).[3] Para que a síntese de fala seja efetuada, Barbosa (2022) aponta a necessidade de reunir e de organizar, de acordo com critérios linguísticos, tecnológicos e de produção de fala, vários elementos, blocos e subestruturas linguísticas e sonoras analisadas previamente. Segundo o autor, a síntese de fala realizada com base no texto escrito consiste no modelo mais empregado nos dias de hoje e incorpora regras linguísticas.

No momento atual, em razão do desenvolvimento de técnicas de aprendizagem da relação estabelecida entre o texto escrito e o som, algumas empresas utilizam sistemas de Aprendizado de Máquina, ainda que sejam "pobres" em termos prosódicos e tenham de percorrer um longo caminho para manipularem o comportamento humano em qualquer situação de comunicação (Barbosa, 2022). A prosódia é entendida, na pesquisa, como os fenômenos cuja unidade descritiva se estende além do nível silábico (Cagliari; Massini-Cagliari, 2003). Esses fenômenos são tradicionalmente agrupados em elementos da

[2] *with a society to perform competitive tasks requiring cognitive processes and communicate with other entities in society by exchanging messages with high information content and shorter representations*.
[3] *production of speech by machines, by way of the automatic phonetization of the sentences to utter*.

melodia da fala, da dinâmica da fala e da qualidade da voz (Cagliari, 1992). O principal desafio dos sistemas de síntese de fala, também conhecidos como *Text-to-Speech* (TTS), é garantir a naturalidade em diferentes contextos, fundamentando-se nos estudos sobre a prosódia das emoções, das atitudes e dos estilos de fala (Barbosa, 2022).[4]

Diante do exposto, a proposta da pesquisa é investigar a entoação (variação melódica da fala) de enunciados declarativos neutros do português brasileiro produzidos, com o auxílio de um recurso de conversão de texto escrito em fala audível, por uma API (*Application Programming Interface*, ou Interface de Programação de Aplicação, em português) elaborada com algoritmos (sequências de comandos matemático-computacionais) de IA da empresa multinacional americana Google.[5] Trata-se, em particular, de uma ferramenta chamada de Google Cloud Text-to-Speech, oferecida pelo Google Cloud Platform e desenvolvida com base na experiência em síntese de fala da DeepMind.[6] Ela afirma oferecer aos usuários benefícios como uma fala de alta fidelidade, uma seleção de voz mais ampla e uma voz exclusiva.[7] Nesse contexto, os enunciados declarativos neutros são aqueles em que o falante (máquina) transmite as informações ao ouvinte (ser humano) sem focalizar (ou enfatizar) qualquer elemento específico da sentença.[8] Em estruturas como essa, além de nenhum constituinte ser destacado individualmente, a atenção é direcionada a todo o enunciado, já que o conteúdo completo da sentença é novo para o interlocutor (Frota, 2000).

A relevância da pesquisa é justificada por certos tópicos. Em primeiro lugar, apesar de a fala sintética na língua portuguesa ser o objeto de estudo de determinadas investigações (Egashira, 1992; Madureira; Silva, C. H.; Aquino, 1995; Aquino, 1997; Gomes, 1998; Barbosa *et al.*, 1999; Ostermann Filho, 2002; Souza, 2010; Sá, 2018; Casanova, 2019; Paixão, 2020; Tunnermann, 2021; Galdino, 2023), observa-se, em plataformas científicas como o Google Scholar, o SciELO e o ResearchGate, a carência de trabalhos nacionais sobre a

[4] A ideia de que os sons se destacam pelo efeito sensorial e expressam sentidos é vista como a base da expressividade da fala (Madureira, 2011, 2016). À vista disso, a pesquisa reconhece a carência de expressividade prosódica em muitos produtos tecnológicos de síntese de fala disponíveis no mercado contemporâneo, mesmo que tenham alcançado avanços significativos no domínio linguístico, em comparação com as ferramentas surgidas nos séculos anteriores.

[5] O serviço Google Cloud Text-to-Speech disponibiliza vozes em português brasileiro e em diversas outras línguas e variedades, o que expande a aplicação em contextos multilíngues e possibilita a adaptação a diferentes realidades linguísticas e culturais. Informações disponíveis em: https://docs.cloud.google.com/text-to-speech/docs/list-voices-and-types?hl=pt-br#list_of_all_supported_languages. Acesso em: 31 de outubro de 2025.

[6] Informações disponíveis em: https://cloud.google.com/text-to-speech?hl=pt-br. Acesso em: 24 de julho de 2023.

[7] Informações disponíveis em: https://cloud.google.com/text-to-speech#benefits. Acesso em: 13 de julho de 2024.

[8] A pesquisa não aborda a focalização prosódica. Para um estudo sobre esse fenômeno discursivo-pragmático com manifestações entoacionais, recomenda-se a leitura dos trabalhos de Gonçalves (1998) e Fernandes (2007a, 2007b).

relação entre a prosódia e a síntese de fala, com o interesse nas características fonológicas da estrutura melódica de vozes geradas por sistemas computacionais de empresas que lideram o setor tecnológico digital. Em segundo lugar, verifica-se a ausência de estudos brasileiros relativos à entoação declarativa neutra produzida pelo Google Cloud Text-to-Speech, segundo uma perspectiva fonológica. Também há o estabelecimento de um diálogo entre a Linguística (Fonética, Fonologia e Linguística Computacional) e outras áreas do conhecimento, como a Ciência da Computação. Além disso, a entoação é responsável, juntamente com o acento e os demais elementos sonoros, por conduzir "o interlocutor na produção da semiose, que é a dedução ou indução dos significados ativados pelo enunciador, da qual nasce o sentido do texto" (Simões, 2009, p. 83), funcionando como o recurso prosódico mais utilizado na caracterização das atitudes do falante (Cagliari, 1992) e na transmissão de significados mesmo quando o contexto discursivo é eliminado e os enunciados aparecem isolados durante os testes perceptivos (Madureira, 2016). Por fim, os resultados do estudo podem oferecer subsídios à discussão acerca da viabilidade de aprimorar a entoação sintética, de modo a torná-la mais acessível e eficaz em diversas situações comunicativas.

A escolha de enunciados declarativos neutros reside no fato de haver uma elevada ocorrência dessas estruturas linguísticas nos mais variados âmbitos de interlocução. Ademais, esses enunciados são recorrentes em respostas a comandos de texto ou de voz feitos a dispositivos eletrônicos que contam com o suporte de assistentes virtuais, frequentemente adotados em uma era de constante progresso tecnológico digital e de automação de processos.[9]

Com base nos apontamentos acima, os objetivos gerais são: (1) caracterizar, de acordo com uma visão integrada entre a Fonologia Entoacional Autossegmental e Métrica (Ladd, 1996, 2008) e a Fonologia Prosódica (Nespor; Vogel, 1986, 2007), auxiliada pela Fonética Acústica, as propriedades entoacionais de enunciados declarativos neutros gerados, no português brasileiro, pelo Google Cloud Text-to-Speech; (2) e estabelecer uma comparação entre a fala sintética e a fala natural em termos de entoação declarativa neutra no português brasileiro. À vista disso, definem-se as seguintes perguntas norteadoras da pesquisa:

[9] A análise de enunciados declarativos neutros é recorrente nos estudos prosódicos (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), pois essas estruturas contêm um contorno entoacional de referência, isento de variações pragmáticas marcadas, o que beneficia a observação dos padrões melódicos particulares de uma língua em uma pesquisa inicial sobre o fenômeno. Essa característica possibilita a identificação dos aspectos gramaticais da entoação e a comparação entre a fala natural e a sintética.

(1) Quais são as características entoacionais de enunciados declarativos neutros gerados, no português brasileiro, pelo Google Cloud Text-to-Speech?
(2) Em que medida a fala sintética do Google Cloud Text-to-Speech reproduz a estrutura fonológica da entoação declarativa neutra observada na fala natural do português brasileiro?

Os objetivos específicos, decorrentes dos objetivos gerais e das perguntas norteadoras da pesquisa, consistem em: (1) identificar os eventos tonais associados ao contorno de entoação dos enunciados declarativos neutros; (2) determinar o padrão entoacional do conjunto de enunciados declarativos neutros; (3) e comparar os aspectos entoacionais dos enunciados sintéticos com os resultados obtidos por Tenani (2002) em uma investigação acerca da estrutura entoacional de enunciados declarativos neutros produzidos, com o auxílio de métodos experimentais, por falantes brasileiros.[10]

Além do caráter pioneiro do trabalho na literatura fonético-fonológica do Brasil, justifica-se a escolha do estudo de Tenani (2002) porque a autora também adota uma visão integrada entre a Fonologia Entoacional Autossegmental e Métrica e a Fonologia Prosódica para caracterizar a estrutura entoacional da declaração neutra no português brasileiro. Além disso, a pesquisa utiliza o *corpus* elaborado pela própria autora para a investigação desse parâmetro melódico. Contudo, a investigação se volta à análise melódica de áudios produzidos por um sistema computacional baseado em IA, constituído de diferentes modelos de voz, o que configura uma abordagem inédita em relação ao estudo de Tenani (2002).

O trabalho se insere em uma linha de pesquisa que une a Fonética, a Fonologia e a Linguística Computacional. Apesar de existirem estudos que analisam a variação melódica em tecnologias de síntese de fala de línguas como o inglês (Pierrehumbert, 1981, 1993), o alemão, o finlandês, o tailandês e o vietnamita (Mixdorff, 2004), além do japonês (Minematsu *et al.*, 2015) e do coreano (Bae *et al.*, 2020), as investigações que associam, de maneira articulada, os critérios de filtragem, de suavização e de interpolação melódica, a notação tonal compatível com a perspectiva teórica da pesquisa e o contraste sistemático entre a fala natural e a fala sintética gerada por um produto comercial de uso massivo ainda são incipientes, tanto

[10] Os objetivos gerais e específicos delineados revelam a intenção da pesquisa de caracterizar e de compreender, de modo sistemático e comparativo, as propriedades entoacionais manifestadas em declarações neutras da fala sintética e da fala natural do português brasileiro. A relevância científica do estudo é demonstrada por meio da integração de abordagens teórico-metodológicas da Fonética e da Fonologia, a fim de viabilizar a investigação, a partir do devido rigor analítico, do desempenho prosódico dos enunciados declarativos neutros e a avaliação da proximidade melódica entre a produção sintética e a natural, com enfoque na referida modalidade enunciativa.

no português brasileiro quanto em outras línguas naturais.[11] No português brasileiro, a tese de Paixão (2020) investiga a prosódia sintética de um *software* destinado à acessibilidade de pessoas com deficiência visual e alcança resultados significativos.[12] A presente pesquisa, entretanto, tem um escopo distinto, na medida em que examina os modelos de fala sintética de um sistema de IA com uma ampla difusão no mercado tecnológico do Brasil e do exterior, aplica uma metodologia original para a avaliação da estrutura fonológica da entoação sintética e privilegia a descrição melódica de construções declarativas neutras documentadas na fala natural pela investigação de Tenani (2002), que é um trabalho consagrado no estudo da prosódia do português brasileiro há mais de duas décadas.[13]

A pesquisa adota uma visão fonológica, orientada por uma abordagem fonética, para a análise da entoação de enunciados declarativos neutros sintetizados, no português brasileiro, pelo Google Cloud Text-to-Speech. Ao comparar os enunciados descritos com as informações encontradas em Tenani (2002) e em outros estudos sobre a entoação do português brasileiro (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), a pesquisa estabelece um diálogo com as investigações anteriores referentes à variação melódica da fala natural e contribui para a compreensão dos padrões entoacionais envolvidos na geração de fala sintética em declarações neutras. Sendo assim, em vez de desconsiderar os trabalhos realizados, a pesquisa pretende colaborar com a literatura disponível, por meio de uma análise detalhada da estrutura melódica de enunciados sintéticos e de como eles se relacionam à entoação declarativa neutra natural.

As hipóteses da pesquisa, relacionadas aos objetivos específicos, são descritas a seguir. Quanto ao primeiro objetivo específico, a hipótese é que sejam identificados os eventos tonais

---

[11] Ainda não há estudos que comparem, no mesmo sistema de síntese de fala e sob a mesma perspectiva teórica, a entoação declarativa neutra de diferentes línguas. Essa ausência reforça o caráter pioneiro da pesquisa, que sistematiza uma metodologia analítica para o português brasileiro e proporciona um referencial para futuros trabalhos comparativos que utilizem a mesma tecnologia ou outros modelos computacionais.

[12] Paixão (2020) realiza testes de percepção com pessoas com deficiência visual, além de análises acústicas e fonológicas de padrões assertivos, interrogativos e continuativos. De acordo com o estudo, as versões de áudio cuja densidade tonal e amplitude de variação melódica são maiores tendem a ser percebidas como mais naturais. Já a presente pesquisa examina as vozes sintéticas de um sistema artificial de ampla utilização na sociedade e aplica, entre outros parâmetros fonéticos e fonológicos, critérios de suavização e de interpolação melódica para a verificação das propriedades exclusivamente gramaticais da entoação declarativa neutra do português brasileiro, em uma perspectiva comparativa com a fala natural.

[13] A ausência de um protocolo comparável, em que o conjunto de vozes, os comandos, os ajustes melódicos e a notação tonal são padronizados, realça a natureza inédita do trabalho e propicia o advento de futuras extensões interlinguísticas no mesmo ambiente computacional, com o controle de variáveis técnicas e fonéticas.

descritos e analisados em trabalhos concernentes à entoação do português brasileiro. Em relação ao segundo objetivo específico, a hipótese é que o padrão entoacional dos enunciados seja definido por uma ascendência tonal no começo da sentença e por uma descendência melódica no término do enunciado, assim como na fala natural. Já no que tange ao terceiro objetivo específico, a hipótese é que os enunciados em análise, embora sejam produzidos por um sistema de síntese de fala, disponham de aspectos entoacionais semelhantes àqueles encontrados nos dados de fala natural descritos por Tenani (2002) e por outros estudos mencionados ao longo da pesquisa (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b).

Essas hipóteses são justificadas pela expansão da aplicação da IA à síntese de fala (Casanova, 2019, 2022; Casanova; Shulby; Aluísio, 2021; Casanova *et al.*, 2021, 2022, 2024). Estudos recentes, como o de van den Oord *et al.* (2016), têm demonstrado que os atuais modelos de áudio, como o WaveNet, disponível no Google Cloud Text-to-Speech, podem produzir uma fala que simula qualquer voz humana e é mais natural do que os principais sistemas de síntese de fala disponíveis na época da pesquisa, com uma diferença percebida em relação à fala natural reduzida em mais da metade dos casos. Além disso, o mesmo modelo pode sintetizar outros tipos de sinais sonoros, como a música, o que reforça a flexibilidade e o potencial das tecnologias de fala atuais para a modelagem de áudio em geral. Esses avanços indicam que a funcionalidade dos sistemas de síntese de fala contemporâneos é capaz de exceder a simulação dos sinais acústicos superficiais e de reproduzir as regularidades linguísticas da fala natural. O uso desses sistemas em serviços digitais atesta, na prática, a relevância e a aceitação social da tecnologia.[14] Diante desse cenário, há um embasamento para a formulação das hipóteses acima sobre a entoação declarativa neutra na fala sintética do português brasileiro gerada pelo Google Cloud Text-to-Speech.[15]

---

[14] O Google Cloud Text-to-Speech é empregado em serviços digitais, o que indica um possível alinhamento entre as saídas acústicas e as expectativas perceptuais dos consumidores. Essa aceitação social sugere que o sistema pode reproduzir uma estrutura fonológica semelhante à da fala natural, com ênfase nos aspectos entoacionais, que contribuem para a inteligibilidade e a compreensão dos enunciados (Pierrehumbert, 1981, 1993). Nesse sentido, a adoção do recurso reforça a premissa de que a naturalidade percebida pode decorrer da correspondência entre a fala sintética e a fala humana, além de legitimar a escolha do sistema para a investigação científica.

[15] Ao adotar as vozes do Google Cloud Text-to-Speech, as hipóteses se fortalecem com a plausibilidade de que, se o sistema é funcional, ele pode gerar, em certa medida, os padrões fonológicos das línguas naturais. Sendo assim, o progresso da IA viabiliza a previsão da reprodução consistente de características entoacionais nos dados sintéticos, mesmo que sejam possíveis eventuais variações técnicas de implementação, sem nenhum prejuízo, a

Ao discutir a vertente teórica criada por Noam Chomsky para o estudo da competência linguística inata,[16] Kato (2001) observa que a teoria da linguagem tem a finalidade de descrever o conhecimento do adulto e de explicar como ele se desenvolve a partir de um estágio inicial. Nessa concepção, os princípios comuns a todas as línguas definem o que é constante em qualquer sistema gramatical, enquanto certas propriedades permanecem abertas para serem especificadas pela experiência. Essa perspectiva é relevante para a pesquisa, que busca compreender até que ponto a IA reproduz, em função de regularidades estatísticas extraídas de dados humanos, os padrões entoacionais e as variações fonológicas inerentes à entoação declarativa neutra do português brasileiro no Google Cloud Text-to-Speech.

O estudo analisa a adequação dos padrões entoacionais descritos para a fala natural quando aplicados à fala sintética no português brasileiro, com foco na declaração neutra. Para tanto, o trabalho utiliza a análise entoacional proposta por Tenani (2002) como referência e avalia os enunciados produzidos pelo Google Cloud Text-to-Speech. Após a comparação dos resultados com a literatura existente sobre a prosódia do português brasileiro (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), o trabalho busca avaliar em que medida o sistema de síntese de fala consegue reproduzir os contornos entoacionais que refletem a organização fonológica da fala natural na declaração neutra. Ao contrário de replicar os resultados anteriores, a investigação se propõe a entender como a prosódia e a síntese de fala se relacionam nesse recorte específico. Os resultados têm implicações para a Linguística, na medida em que evidenciam como as regularidades entoacionais da gramática do português brasileiro se manifestam, para os enunciados declarativos neutros, em contextos de síntese de fala. Ademais, os resultados são relevantes para o aprimoramento tecnológico de ferramentas de fala sintética, visto que os desenvolvedores desses sistemas devem conhecer, em um processo colaborativo com uma equipe de linguistas, as características fonológicas básicas do português brasileiro e a forma como elas têm sido reproduzidas pela síntese de fala moderna, para que possam propor soluções eficientes voltadas à melhoria do processamento computacional de fala.

princípio, para as categorias simbólicas.

[16] Os princípios teóricos defendidos pela vertente teórica em questão podem ser encontrados em trabalhos clássicos de Chomsky (1957, 1965, 1986, 1993).

Apesar de haver trabalhos que estabeleçam uma separação estrita entre a Fonética e a Fonologia, a pesquisa adota uma perspectiva de diálogo entre as duas disciplinas, que se articulam de maneira complementar. A manipulação responsável de medidas numéricas do correlato acústico da entoação possibilita a recuperação de categorias tonais que caracterizam a gramática entoacional de enunciados declarativos neutros no português brasileiro. Essa estratégia transita entre o nível simbólico, representado por eventos tonais, e o nível numérico, representado por trajetórias contínuas da melodia. A combinação entre as pistas acústicas e a interpretação de categorias tonais formaliza os contornos melódicos e evidencia que a descrição da entoação declarativa neutra obtém benefícios da ação cooperativa entre a Fonética e a Fonologia. Tal ponto de vista defende uma concepção de que as duas disciplinas não são dicotômicas, mas, na verdade, correlacionadas no tratamento de fenômenos como a entoação.

A análise acústica se concentra em aspectos específicos e revela que as pistas fonéticas observadas também veiculam informações gramaticais. Esses elementos propiciam a identificação de categorias melódicas e indicam que a informação gramatical não se restringe somente ao nível simbólico, mas também se manifesta por meio de expedientes acústicos. Nesse quadro, a Fonética fornece o correlato mensurável que sustenta a informação fonológica e estabelece uma continuidade de mapeamento entre as pistas físicas e as categorias fonológicas de interpretação. As curvas entoacionais analisadas mostram que a Fonética ultrapassa o caráter meramente físico e participa da codificação da estrutura fonológica. A Fonética e a Fonologia operam, portanto, de forma integrada: a primeira proporciona a base acústica e a segunda a formaliza em categorias simbólicas, em consonância com enfoques que se aproximam da Fonologia de Laboratório (Beckman; Kingston, 1990; Pierrehumbert; Beckman; Ladd, 2000; Albano, 2017), que parte da evidência empírica da fala para a condução da análise fonológica. Essas duas disciplinas, juntas, compõem o cerne explicativo da entoação declarativa neutra do português brasileiro na fala sintética gerada pelo Google Cloud Text-to-Speech.

A pesquisa adquire um caráter fonético ao empregar, além da inspeção auditiva, os procedimentos de extração e de registro da variação melódica da fala, como a filtragem e a suavização das curvas entoacionais e a interpolação de pontos tonais ausentes.[17] Esses procedimentos garantem que a investigação se baseie em pistas fonéticas contrastivas, e não em oscilações acústicas sem qualquer explicação linguística, para fundamentar a interpretação

[17] O procedimento adotado é conhecido como “inspeção auditiva”, porque se trata de uma observação controlada pelo pesquisador, diferente da “análise auditiva”, que, em geral, pressupõe o julgamento de ouvintes externos.

fonológica. Para esclarecer ainda mais o componente fonético do estudo, são realizadas verificações pontuais de picos e de vales melódicos, que podem subsidiar a identificação de diferentes aspectos tonais. Assim, torna-se explícita a forma como os detalhes acústicos orientam a decisão sobre os elementos melódicos e ajudam a manter o enfoque nas propriedades estritamente gramaticais da entoação declarativa neutra na fala sintética do português brasileiro.

A investigação se insere em um cenário brasileiro e internacional em que as diferentes línguas naturais podem ser analisadas sob uma perspectiva de interface entre a prosódia e a síntese de fala. O estudo da entoação de enunciados declarativos neutros é uma contribuição valiosa para a compreensão de como esses padrões melódicos são reproduzidos em modelos de síntese de fala, com a consequente promoção de benefícios práticos relacionados, por exemplo, à acessibilidade e à inclusão. Ao articular a Linguística e a Ciência da Computação, com a possibilidade de extensão à Engenharia Eletrônica e à Engenharia Elétrica, a investigação supera o caráter descritivo e se configura como uma pesquisa interdisciplinar voltada à promoção de avanços científicos, tecnológicos e sociais.

A estrutura do trabalho é organizada em seções que abrangem os aspectos fundamentais da investigação sobre a entoação declarativa neutra do português brasileiro na fala sintética do Google Cloud Text-to-Speech. Após a introdução, em que são descritos o contexto, a proposta, os objetivos e as hipóteses da pesquisa, a segunda seção é dedicada à fundamentação teórica, em que são expostos os principais conceitos e teorias que embasam o estudo. Essa seção é organizada em três partes: a primeira subseção aborda os pressupostos conceituais utilizados na análise das propriedades fonético-fonológicas da variação melódica da fala; a segunda explora a entoação do português brasileiro, com foco no domínio prosódico de associação tonal e nos aspectos fonético-fonológicos de enunciados declarativos neutros; e a terceira delineia os conceitos básicos da Linguística Computacional que respaldam o trabalho e apresenta um panorama da síntese de fala, com destaque aos aspectos históricos, aos principais métodos e às abordagens utilizadas na produção entoacional.

A terceira seção, referente à metodologia, descreve a amostra composta de enunciados declarativos neutros extraídos do *corpus* elaborado por Tenani (2002) e a abordagem utilizada para a investigação da entoação da fala sintética. A seção é dividida em duas partes: a primeira trata dos materiais utilizados, com a descrição dos modelos de voz disponíveis no Google Cloud Text-to-Speech e das características gramaticais (fonológicas, morfológicas e sintáticas) da amostra estudada. A segunda parte descreve os métodos adotados, que incluem as etapas de descrição da estrutura entoacional dos arquivos de fala sintética, como a

identificação dos eventos tonais, a determinação do padrão entoacional e a comparação com os resultados de Tenani (2002) e de outras pesquisas sobre a entoação do português brasileiro.

A quarta seção, concernente aos resultados e às discussões, sistematiza os achados da pesquisa, organizados em três subseções. A primeira subseção aborda os acentos tonais, a segunda foca no tom de fronteira e a terceira examina as configurações entoacionais dos enunciados declarativos neutros gerados pelo Google Cloud Text-to-Speech. A pesquisa é respaldada na Fonética Acústica, na Fonologia Entoacional Autossegmental e Métrica e na Fonologia Prosódica, o que permite a verificação tanto dos aspectos físicos relevantes para o sistema linguístico quanto das propriedades fonológicas da entoação declarativa neutra. A análise acústica é utilizada como uma ferramenta para a identificação das propriedades fonológicas da entoação sintética, sem a necessidade da realização de cálculos físico-matemáticos aprofundados. Durante a seção, são feitas comparações entre os aspectos entoacionais dos enunciados sintéticos e os resultados de Tenani (2002), bem como um diálogo com outros estudos relativos à estrutura melódica do português brasileiro.

Por fim, a última seção, destinada à conclusão, reúne, antes das referências,[18] os principais resultados do estudo e destaca as contribuições da pesquisa para o campo da prosódia da fala sintética. São discutidas as implicações dos achados e apresentadas algumas sugestões para futuras investigações, com ênfase na melhoria da modelagem entoacional em sistemas de síntese de fala e no aprofundamento da caracterização das propriedades melódicas da fala sintética no contexto da variedade brasileira do português.

[18] O pesquisador, apesar de se responsabilizar pelo trabalho, agradece aos colegas André Luiz Machado e Carlos Elísio Nascimento da Silva a contribuição em aspectos relacionados às referências.

2 FUNDAMENTAÇÃO TEÓRICA

Esta seção é referente à exposição da fundamentação teórica e organizada em três partes. A primeira subseção aborda os pressupostos conceituais utilizados na análise das propriedades fonético-fonológicas da variação melódica da fala sintética. A segunda subseção, por sua vez, relaciona-se à entoação do português brasileiro, com foco no domínio prosódico de associação tonal e nos aspectos fonético-fonológicos de enunciados declarativos neutros. Já a terceira subseção delineia os fundamentos básicos da Linguística Computacional que embasam a pesquisa, bem como apresenta um panorama da síntese de fala, com ênfase nos aspectos históricos, nos principais métodos e nas abordagens de produção entoacional.

2.1 ANÁLISE DA ENTOAÇÃO: FONÉTICA ACÚSTICA, FONOLOGIA ENTOACIONAL E FONOLOGIA PROSÓDICA

De modo geral, a prosódia, que consiste em um dos eixos de conhecimento ao qual a pesquisa pertence, pode ser investigada, na Linguística, tanto do ponto de vista fonético quanto fonológico. Tradicionalmente, a Fonética é a disciplina que estuda os sons da língua como entidades passíveis de descrição a partir de suas características físicas e articulatórias (Abaurre, 1993). Já a Fonologia é a disciplina dedicada ao estudo do valor que as línguas naturais atribuem aos sons selecionados dentre um inventário universal de possibilidades de articulação (Abaurre, 2010). Além disso, o componente fonológico da gramática pode ser compreendido como um sistema de regras, de parâmetros e de normas que estruturam as oposições sonoras da língua, bem como as alternativas de realização dessas oposições utilizadas pelos falantes em contextos linguísticos e extralinguísticos (Abaurre, 2003, 2013).[19]

Apesar da relevância da prosódia para os estudos sonoros (fonéticos e fonológicos), não se descarta a possibilidade de ela estabelecer um diálogo com os outros componentes linguísticos, dos mais formais aos mais discursivos (Cavalcante; Scarpa, 2022). Ademais, a prosódia se relaciona com a música, haja vista que esta, assim como a fala, dispõe de melodia,

[19] A pesquisa adota uma perspectiva de interface entre a Fonética e a Fonologia, em que a análise fonético-acústica dos contornos entoacionais objetiva identificar a existência de regularidades sistemáticas passíveis de interpretação como categorias prosódicas portadoras de algum valor gramatical. Essa abordagem viabiliza a descrição precisa da entoação na fala sintética e a comparação com os padrões fonológicos observados na fala natural. Os dados fonético-acústicos da fala sintética são essenciais para a identificação de categorias fonológicas da estrutura entoacional, o que exige uma análise capaz de considerar tanto o aspecto físico do sinal acústico quanto a estrutura simbólica e sistemática da fala. Desse modo, pode-se avaliar se os padrões melódicos produzidos pelo sistema de IA correspondem às estratégias prosódicas descritas para as declarações neutras na fala natural.

como o tom e a entoação, e de harmonia e de pulsação, como o acento, o ritmo e a duração (Massini-Cagliari; Cagliari, 2012; Massini-Cagliari, 2015). Os aspectos gerais do campo são apresentados a seguir.[20]

> O termo recobre, nos estudos lingüísticos, uma gama variada de fenômenos que abarcam os parâmetros de altura, intensidade, duração, pausa, velocidade de fala, bem como o estudo dos sistemas de tom, entoação, acento e ritmo das línguas naturais, gama esta que demanda enfoques especializados aos fatos considerados. Dois grandes pólos de interesse nos estudos prosódicos podem ser traçados. O primeiro deles é o tratamento acústico, mensurável, instrumental de altura, intensidade e quantidade, correlatos perceptuais de freqüência, volume e duração. A atenção tem sido voltada não só para o estudo dos parâmetros individuais em si e sua relação com os demais fenômenos fônicos segmentais, adicionados a considerações sobre velocidade de fala e qualidades da voz, como também aos correlatos acústicos dos sistemas de acento, ritmo e entoação nas línguas. O segundo pólo de interesse é a consideração fonológica das organizações e representações dos sistemas de acento, ritmo e entoação nas línguas e suas interfaces com os demais componentes lingüísticos. Portanto podem-se esperar interseções e/ou gradações no tratamento teórico e metodológico entre os objetos de interesse de cada um destes pólos – do "mais fonético" ao "mais fonológico" (Scarpa, 1999, p. 8).

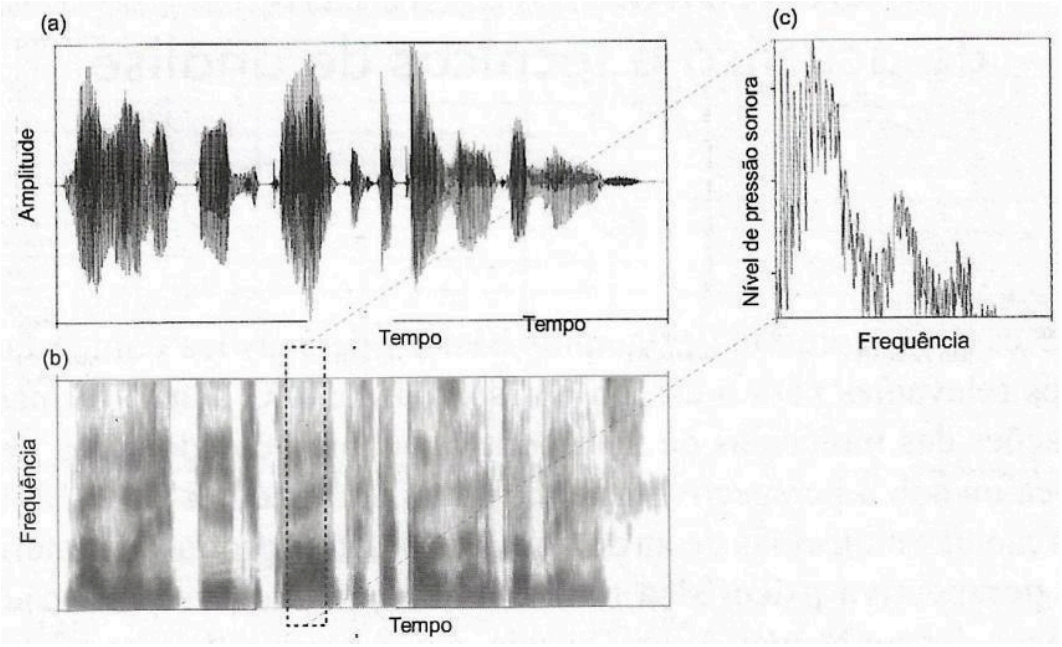
A primeira disciplina dos estudos prosódicos com a qual a pesquisa dialoga é a Fonética Acústica,[21] cujo objetivo consiste em estudar as "propriedades físicas dos sons da fala a partir de sua transmissão do falante ao ouvinte" (Cristófaro Silva, 2022, p. 23, grifos no original). A análise acústica do sinal de fala abrange três tipos de representações gráficas: a forma de onda, que relaciona o tempo à amplitude; o espectrograma, que relaciona o tempo à frequência; e o espectro, que relaciona a frequência à amplitude (Cristófaro Silva *et al.*, 2019). O tempo corresponde à duração da produção sonora, a amplitude se refere à energia gasta na produção sonora e a frequência equivale ao número de repetições de uma onda sonora por unidade temporal (Cristófaro Silva *et al.*, 2019). Essas representações gráficas são visualizadas na Figura 1.

---

[20] Na citação direta, há uma inversão de conceitos teóricos, já que a frequência é, na verdade, um parâmetro acústico, cujo correlato perceptual, no estudo da entoação, é a altura.

[21] No decorrer da subseção, os conceitos fonéticos, embora discutidos nos estudos citados (por exemplo, Massini-Cagliari, 1992; Cagliari, 1992, 2012a; Abaurre, 1998; Scarpa, 1988; Cagliari; Massini-Cagliari, 2003; Massini-Cagliari; Cagliari, 2012; Barbosa, 2019; Cristófaro Silva *et al.*, 2019; Cristófaro Silva, 2022; Berti *et al.*, 2023, 2025), têm origem, em parte, em trabalhos clássicos da Fonética, como os de Pike (1945), Ladefoged (1962), Halliday (1963, 1970), Abercrombie (1967), Bolinger (1972), Cagliari (1981), Cruttenden (1986), Couper-Kuhlen (1986) e Laver (1994). Essas obras, que orientam investigações acústicas sobre o português e outras línguas, estabelecem as bases teóricas para a compreensão da relação entre as propriedades acústicas dos sons e os efeitos auditivos relacionados à entoação, bem como aos demais elementos prosódicos e segmentais.

Figura 1 – Representação da forma de onda (a), do espectrograma (b) e do espectro (c)



Fonte: Cristófaro Silva *et al.* (2019, p. 38).

As investigações em Fonética Acústica, de forma geral, preocupam-se com a estrutura física dos sons da fala, a fala sintética e o reconhecimento automático de fala (Massini-Cagliari; Cagliari, 2012). Trata-se, portanto, de uma disciplina que, além de favorecer a Linguística, contribui para o desenvolvimento de tecnologias em que são utilizados os elementos sonoros da fala, como na Engenharia de Telecomunicações, principalmente no contexto da telefonia, e na Ciência da Computação, com enfoque nos programas destinados à produção e ao reconhecimento de fala (Massini-Cagliari; Cagliari, 2012). Soma-se ao descrito a ampliação progressiva das contribuições da Linguística, em uma perspectiva ampla, para o setor de IA, impulsionada pelas iniciativas de combate ao negacionismo científico, com implicações para o desenvolvimento de sistemas de percepção e de síntese de fala (Sene; Massini-Cagliari, 2023).

O diálogo da pesquisa com a disciplina de Fonética Acústica se justifica pelo fato de que são analisadas as configurações fonéticas, com algum valor linguístico, das proeminências tonais e da direção da curva melódica, que refletem os padrões fonológicos do português brasileiro. Esse argumento indica que o uso de um *software* para o tratamento acústico dos dados de fala, sem uma interpretação científica do fenômeno, não é suficiente para justificar o enquadramento de uma pesquisa nos estudos fonéticos (Silva, A., 2010, 2020a).

Além da Fonética Acústica, cujos métodos e ferramentas são detalhados na seção de

metodologia, é adotada, com a intenção de descrever e de analisar a estrutura entoacional dos enunciados declarativos neutros sintéticos, uma visão integrada entre a Fonologia Entoacional Autossegmental e Métrica (ou apenas Fonologia Entoacional), conforme a proposta elaborada por Ladd (1996, 2008) com base em trabalhos como os de Pierrehumbert (1980), Beckman e Pierrehumbert (1986) e Pierrehumbert e Beckman (1988), e a Fonologia Prosódica, desenvolvida por Nespor e Vogel (1986, 2007).[22] Tais arcabouços teóricos são uma reação à Fonologia Gerativa Padrão, apresentada na obra *The Sound Pattern of English*, escrita por Chomsky e Halle (1968).[23] Essa obra não fornece nenhum tipo de formalismo à representação e à manipulação de propriedades prosódicas como a altura (*pitch*) e a duração (Abaurre; Wetzels, 1992), o que motiva a proposta de outras teorias para a investigação dos elementos prosódicos, as quais preservam os princípios teóricos básicos da Fonologia Gerativa Padrão.

Na literatura fonético-fonológica internacional, uma visão integrada entre a Fonologia Entoacional Autossegmental e Métrica e a Fonologia Prosódica é adotada, em um primeiro momento, por estudos como os de Hayes e Lahiri (1991), Jun (1996) e Frota (2000). Para a investigação da prosódia do português brasileiro, a relação entre esses dois modelos é adotada, a título de exemplo, por Frota e Vigário (2000), Tenani (2002), Fernandes (2007b) e Serra (2009), que trabalham com dados de fala natural, e por Paixão (2020), que lida com a fala sintética voltada, no escopo das tecnologias inclusivas, às pessoas com deficiência visual. Além de tais trabalhos, são produzidos outros estudos acerca do assunto, como os de Truckenbrodt, Sandalo e Abaurre (2009), Vigário e Fernandes-Svartman (2010), Silvestre e Cunha (2013), Toneli (2014), Toneli, Vigário e Abaurre (2014), Frota *et al.* (2015), Serra (2016), Soncin e Tenani (2016), Fernandes-Svartman e Romano (2017), Silvestre (2017) e Toneli, Abaurre e Vigário (2018).

---

[22] Selkirk (1984) propõe o modelo *end-based* para a análise dos constituintes prosódicos, segundo o qual as fronteiras sintáticas servem de referência para a delimitação desses domínios fonológicos (Tenani, 2017). No entanto, o trabalho utiliza o modelo *relation-based*, formulado por Nespor e Vogel (1986, 2007), que considera as relações sintáticas como informações relevantes para a configuração dos constituintes prosódicos (Tenani, 2017), em consonância com estudos acerca do português brasileiro, como os de Tenani (2002), Fernandes (2007b), Serra (2009) e Fernandes-Svartman (2024a, 2024b).

[23] A escolha dos quadros teóricos da Fonologia Entoacional Autossegmental e Métrica e da Fonologia Prosódica decorre dos objetivos da pesquisa, que examina a entoação declarativa neutra do português brasileiro produzida por um sistema de IA. A Fonologia Entoacional Autossegmental e Métrica, proposta por Ladd (1996, 2008) e inspirada em Pierrehumbert (1980), permite a representação da estrutura melódica e a descrição do contorno entoacional dos enunciados. Já a Fonologia Prosódica, desenvolvida por Nespor e Vogel (1986, 2007), define os domínios prosódicos hierárquicos e esclarece a forma de organização dos constituintes fonológicos. A combinação dessas abordagens viabiliza o entendimento da entoação sintética em relação à estrutura fonológica da língua e segue a tradição de estudos sobre o português brasileiro (Frota; Vigário, 2000; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Serra, 2009; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). A Fonética Acústica, por sua vez, estabelece as bases instrumentais que tornam possível a observação empírica da correspondência entre as propriedades físicas da variação melódica da fala e as categorias fonológicas analisadas.

Como mencionado previamente, a Fonologia Entoacional Autossegmental e Métrica é formulada dentro do escopo da Gramática Gerativa.[24] Essa concepção de gramática procura esclarecer o conhecimento linguístico ou a faculdade da linguagem dos falantes-ouvintes de qualquer língua, sob a premissa de que a gramática constitui um conjunto de regras que define a relação entre os sons e os significados em uma língua (Lee, 2017).

Kato (1997) argumenta que, na perspectiva gerativista, a língua é considerada uma faculdade interna de natureza individual, baseada em uma gramática bioprogramada (Chomsky, 1957, 1965, 1986, 1993). De acordo com a autora, o foco teórico é o conhecimento inconsciente do falante acerca da língua, ao invés do desempenho externo. Esse conhecimento resulta do funcionamento automático de um sistema computacional responsável pela estruturação da gramática linguística, e não da simples internalização de hábitos.[25]

Toda teoria de aquisição da linguagem pressupõe uma concepção de língua, que separa a visão externa, entendida como um conjunto de enunciados observáveis, da visão interna, correspondente ao conhecimento individual, conforme observado por Kato (2002). Segundo a autora, a segunda perspectiva, defendida por Chomsky (1986), concebe a gramática inata como uma capacidade cognitiva compartilhada pela espécie humana (Chomsky, 1965), constituída de princípios fixos e de opções de variação determinadas pelo *input* recebido.[26]

De acordo com a Fonologia Entoacional Autossegmental e Métrica, a entoação apresenta uma organização fonológica própria.[27] Nesse modelo teórico, o contorno entoacional é constituído de uma sequência linear de eventos tonais discretos,[28] associados a

---

[24] Na subseção, todos os conceitos, princípios e definições pertencentes à Fonologia Entoacional Autossegmental e Métrica são redigidos a partir do trabalho de Ladd (2008), que, por sua vez, fundamenta-se especialmente na proposta de análise de Pierrehumbert (1980) para a entoação da língua inglesa. Quando são utilizadas pesquisas e adaptações de outros autores, elas são referenciadas nos parágrafos correspondentes.

[25] A pesquisa analisa a entoação declarativa neutra do português brasileiro produzida pelo Google Cloud Text-to-Speech e examina em que medida os contornos melódicos sintéticos reproduzem as regularidades prosódicas encontradas na fala humana, decorrentes de uma gramática fonológica orientadora da produção linguística de falantes proficientes, cujas amostras de áudio compõem o banco de dados utilizado para o treinamento dos modelos de síntese de fala.

[26] Nos estudos linguístico-computacionais, essa distinção é crucial, pois evidencia que a análise da entoação sintética deve considerar, para fins comparativos, as propriedades prosódicas derivadas da estrutura fonológica da língua natural, refletidas nos dados de fala humana utilizados na criação de modelos de síntese de fala. Dessa forma, defende-se que os sistemas de IA processam os padrões estatísticos provenientes dos dados de fala humana, em que se encontram as regularidades compatíveis com a estrutura fonológica da língua natural.

[27] Sugere-se ao leitor consultar o *Dicionário Pedagógico de Fonologia Entoacional do Português Brasileiro* (Tojeira-Ramos, 2025), produzido em formato digital, caso tenha dúvidas quanto à compreensão dos conceitos teóricos. Disponível em: https://www.lexonomy.eu/35f8qhdy. Acesso em: 10 de março de 2025.

[28] Massini-Cagliari e Cagliari (2012) afirmam que o primeiro processo de produção da fala humana é de natureza neurolinguística, caracterizado pela associação de ideias (ou seja, de abstrações) aos sons correspondentes ao que se pretende expressar, segundo a ordem e as regras linguísticas. À vista disso, o estudo da entoação da fala sintética não é limitado à avaliação da fidelidade melódica na reprodução computacional, pois também considera, ainda que de modo indireto, o mecanismo neurolinguístico subjacente à fala humana, da qual são

pontos específicos da cadeia segmental.[29] A teoria distingue os eventos tonais das transições na estrutura tonal, reconhece certas partes como relevantes do ponto de vista linguístico e classifica outras apenas como as transições que preenchem a variação tonal entre os eventos locais (Lucente, 2014). O motivo pelo qual o modelo é autossegmental e métrico decorre das razões expostas em seguida.

> O modelo é autossegmental porque tem camadas separadas para segmentos (vogais e consoantes) e tons (H,L). É métrico porque assume que os elementos nessas camadas estão contidos em um conjunto hierarquicamente organizado de constituintes fonológicos (Gussenhoven, 2002, p. 27, tradução nossa).[30]

De uma perspectiva fonético-acústica, o correlato físico de uma sequência de eventos tonais é a frequência fundamental (F_0), normalmente mensurada em Hertz (Hz) e responsável pela produção do efeito auditivo de altura sonora (Massini-Cagliari; Cagliari, 2012), ou seja, a percepção de um som como grave ou agudo em um trecho que corresponde a uma unidade linguística ou enunciado (Barbosa, 2019). Em outras palavras, as variações melódicas da fala, exibidas na Figura 2, são associadas à variação acústica da F_0 (Massini-Cagliari, 1992), que corresponde à quantidade de ciclos por segundo executados pelas pregas vocais (Berti *et al.*, 2025).

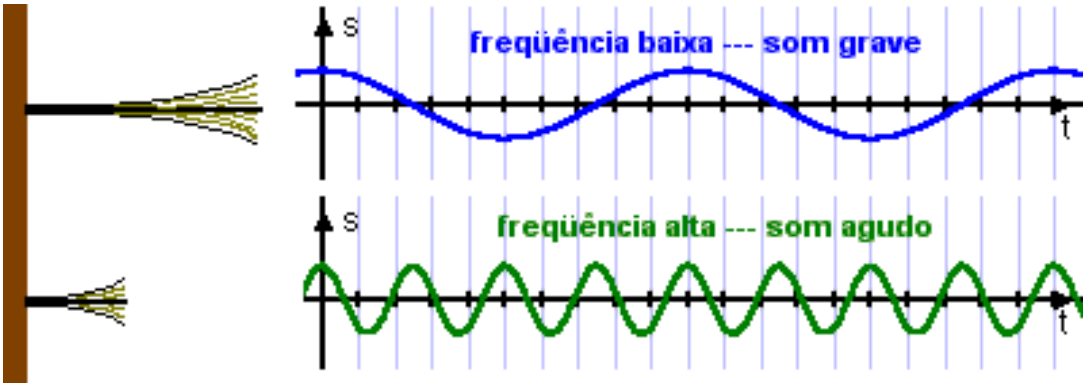
---

provenientes os dados necessários para o desenvolvimento dos modelos de síntese de fala.

[29] Ao se fundamentar nos trabalhos de Chomsky (1986, 1993), Kato (1995, 1997) esclarece que a Língua-I é concebida como um sistema computacional interno e inconsciente, distinto da Língua-E, entendida como um produto externo do desempenho. Essa distinção evidencia a necessidade da investigação da entoação não apenas como um fenômeno físico, mas também como uma manifestação do conhecimento gramatical dos falantes proficientes, cujos dados naturais "alimentam" os sistemas de síntese de fala. A entoação produzida por modelos computacionais possibilita a observação das regularidades decorrentes de uma gramática fonológica humana, visto que os padrões contidos nos dados representam as características estruturais do conhecimento linguístico dos falantes. Embora esse conhecimento interno organize, segundo Chomsky (1993), a relação entre a forma fonética e a forma lógica, ele não é encontrado nas máquinas, pois os computadores não dispõem de uma competência gramatical inata. Os contornos entoacionais produzidos computacionalmente se ancoram nas regularidades encontradas nos dados humanos utilizados no treinamento estatístico e fazem com que seja possível uma análise indireta das restrições fonológicas que orientam o desempenho dos falantes.

[30] *The model is autosegmental because it has separate tiers for segments (vowels and consonants) and tones (H,L). It is metrical because it assumes that the elements in these tiers are contained in a hierachically organized set of phonological constituents.*

Figura 2 – Representação de uma frequência baixa e de uma frequência alta



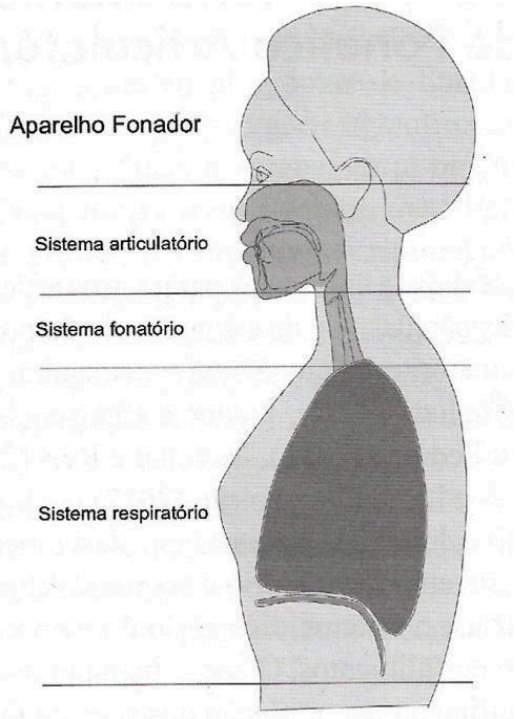
Fonte: Oliveira, L., Veit e Schneider (2002).

Conforme a abordagem teórica adotada, o principal parâmetro acústico verificado em uma análise entoacional é a F_0. Entretanto, apesar de se atentar à variação melódica da fala, não é possível separá-la, de forma integral, da duração (Brigner, 1988; Beerends, 1989; Cumming, 2010; Steffman; Jun, 2019), que, por sua vez, relaciona-se ao ritmo linguístico (Massini, 1991; Massini-Cagliari, 1992, 1995, 1999). Nesse sentido, ao investigar os aspectos fonológicos da entoação de uma língua, o pesquisador não deve ignorar a relação entre os padrões rítmicos e entoacionais dos enunciados (Halliday, 1963, 1970; Cagliari, 2007, 2012a; Massini-Cagliari; Cagliari, 2012).

Os sons periódicos da fala são formados por harmônicos, que correspondem a múltiplos inteiros da primeira frequência (Massini-Cagliari; Cagliari, 2012). A F_0 é caracterizada, na fala natural, por vibrações das pregas vocais (fonação) capazes de produzir, na corrente de ar, uma forma de onda acústica periódica (Cagliari, 2012a). Tais processos são distintos da fala sintética baseada em IA com a qual a pesquisa trabalha, em que a F_0 é gerada por meio de mecanismos linguístico-computacionais.[31] Para o entendimento do processo natural de fonação, a Figura 3 ilustra uma representação esquemática dos sistemas respiratório, fonatório e articulatório envolvidos na geração dos sons da fala.

[31] Kato (1999) enfatiza que, sob a perspectiva gerativista, a gramática deve ser entendida como um módulo mental específico, geneticamente programado (Chomsky, 1986), e não como um mecanismo multifuncional. Trata-se de uma observação que fortalece a premissa da pesquisa de que os fenômenos entoacionais são processados por mecanismos linguístico-cognitivos especializados, que, em virtude do contínuo progresso da tecnologia digital, podem ser simulados e reproduzidos a partir de modelos atuais de síntese de fala.

Figura 3 – Representação dos sistemas respiratório, fonatório e articulatório no aparelho fonador



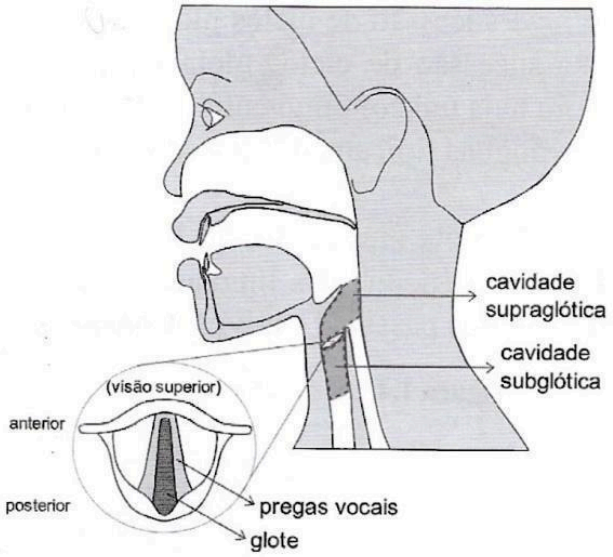
Fonte: Cristófaro Silva *et al.* (2019, p. 12).

Na fala natural, ao ajustar a laringe, em que se localizam as pregas vocais, um falante pode diminuir ou aumentar a F_0 em relação às faixas de frequências mais altas e mais baixas que costuma usar na fala normal, com fins expressivos, como a expressão de fúria, de raiva e de desespero (Cagliari; Massini-Cagliari, 2003). As Figuras 4 e 5 ilustram, nessa ordem, os órgãos fonatórios e o processo de fonação baseado nos movimentos de abertura e de fechamento da glote, definida como a passagem formada entre as pregas vocais (Massini-Cagliari; Cagliari, 2012).[32] Esses ajustes das estruturas laríngeas, em cooperação com o desempenho de outros órgãos, são necessários para que a entoação seja produzida, pois permitem a realização de variações melódicas que transmitem informações semânticas e pragmáticas no decorrer da comunicação oral.[33]

[32] Não há um sistema fisiológico exclusivo para a fala, já que a produção vocal depende da cooperação de órgãos utilizados também em outros processos vitais. Assim, a fala é o resultado da adaptação e da coordenação desses órgãos, cuja ação conjunta possibilita a produção dos sons envolvidos na comunicação oral (Callou; Leite, 1994; Cagliari, 2007; Massini-Cagliari; Cagliari, 2012; Cristófaro Silva, 2002).

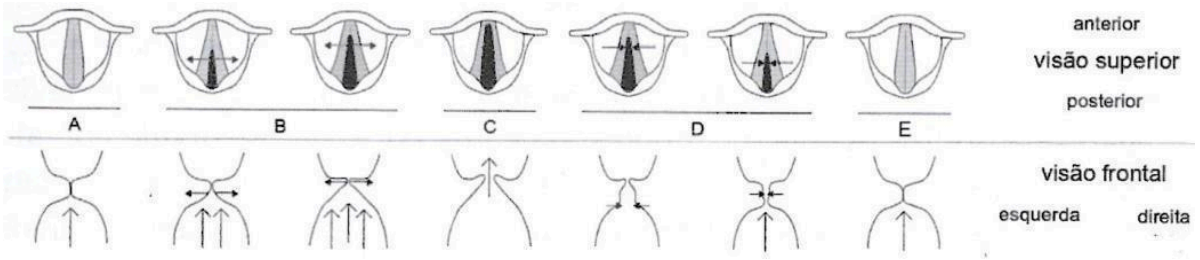
[33] A fala emerge da coordenação temporal dos órgãos respiratórios, laríngeos e supralaríngeos, cuja sincronia permite implementar, por exemplo, os contornos entoacionais (Callou; Leite, 1994; Cagliari, 2007; Massini-Cagliari; Cagliari, 2012; Cristófaro Silva, 2002).

Figura 4 – Representação dos órgãos fonatórios com destaque às cavidades supraglóticas e subglóticas



Fonte: Cristófaro Silva *et al.* (2019, p. 13).

Figura 5 – Representação do processo de fonação



Fonte: Cristófaro Silva *et al.* (2019, p. 12).

Das distinções referentes aos valores da F_0, que também se relacionam aos aspectos anatomofisiológicos das pregas vocais (Berti *et al.*, 2023), resultam os padrões de direção da curva de altura, tais como ascendente, descendente, ascendente-descendente, entre outros (Abaurre, 1998). Já o âmbito de altura (ou tessitura) corresponde aos diferentes níveis de altura em um contorno entoacional (Abaurre, 1998), ou seja, diz respeito à "gradação entre o limite mais alto e o mais baixo no espectro de altura" (Scarpa, 1988, p. 68). O contorno entoacional, no que lhe diz respeito, faz referência "ao formato, à configuração descritiva e quase visual do enunciado em termos de tessitura e de direção da curva" (Scarpa, 1988, p. 68). No nível fonético, um contorno descendente pode ser realizado de diferentes formas, como de médio a baixo ou de alto a baixo, do mesmo modo que um contorno ascendente pode ser realizado de médio a alto ou de baixo a alto (Abaurre, 1998).

Em línguas tonais, as sílabas dos itens lexicais dispõem de uma altura melódica

determinada (Massini-Cagliari; Cagliari, 2012), de modo que o papel primordial dos tons é diferenciar os significados lexicalizados (Cagliari, 1992). Por outro lado, em línguas entoacionais, os variados tipos de enunciados apresentam os padrões melódicos previamente estabelecidos pelo sistema linguístico (Massini-Cagliari; Cagliari, 2012; Massini-Cagliari, 2017).

Tipologicamente, o português é classificado como uma língua entoacional (Massini-Cagliari; Cagliari, 2012), em que os eventos tonais não têm a função de distinguir o sentido das palavras (significado semântico), como ocorre nas línguas tonais (Cagliari, 1992), mas a de veicular, por exemplo, o sentido e a função comunicativa dos enunciados (significado pragmático). Desse modo, no português, as variações tonais dispõem de uma função distintiva no nível da frase, já que diferenciam as frases declarativas das interrogativas por meio dos padrões entoacionais (Callou; Leite, 1994). Enquanto as frases declarativas apresentam um padrão descendente, as frases interrogativas normalmente têm um padrão ascendente, conforme os exemplos ilustrados no Quadro 1, reproduzido por Massini-Cagliari (2017), a partir de uma adaptação de Massini-Cagliari e Cagliari (2001). Evidencia-se, portanto, que os aspectos de produção e de percepção da prosódia da fala exercem um efeito direto na comunicação, pois desempenham, entre outras funções linguísticas, o papel de indicar se um enunciado é uma declaração ou uma interrogação (Constantini; Barbosa, 2015).[34]

Quadro 1 – Padrão entoacional das frases declarativas e interrogativas no português

Significado	**Exemplo**
declaração, asserção	 Ontem choveu muito.
interrogação	 Ontem choveu muito?

Fonte: Reprodução de Massini-Cagliari (2017, p. 26), com base em Massini-Cagliari e Cagliari (2001, p. 118).

Em línguas entoacionais, são distinguidos, segundo a Fonologia Entoacional Autossegmental e Métrica, dois tipos básicos de eventos tonais para a caracterização da estrutura entoacional dos enunciados. Esses eventos são conhecidos como acentos tonais

[34] É importante observar que, a depender da classe dos enunciados interrogativos e da variedade dialetal investigada, o padrão melódico pode não exibir a configuração terminal ascendente (Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Fernandes-Svartman, 2024a, 2024b).

(*pitch accents*) e tons associados a fronteiras prosódicas (*edge tones*), que podem ser formados por apenas dois níveis de tons primitivos (*primitive level tones*) ou alvos de altura (*pitch targets*): alto (H – *high*) e baixo (L – *low*).

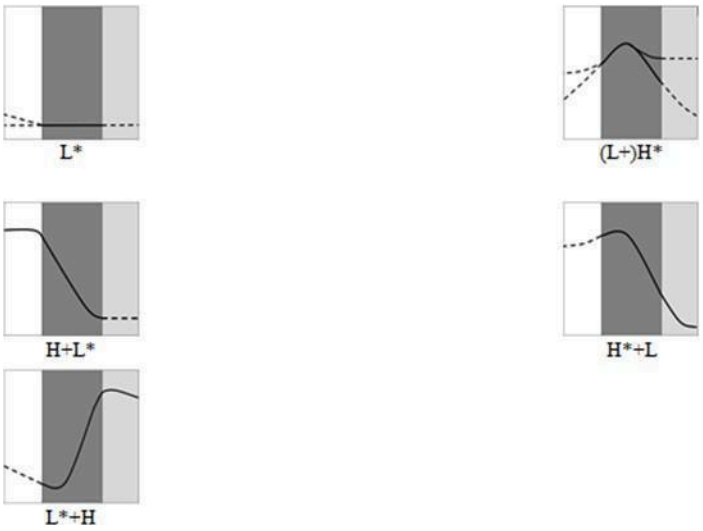
As especificidades dos eventos tonais são descritas a seguir. No entanto, antes da explicação sobre cada um deles, deve-se alertar o leitor sobre a premissa de que as notações que descrevem os contornos entoacionais específicos de uma língua ou dialeto podem apresentar distintas realizações fonéticas em outras línguas ou dialetos, uma vez que os eventos tonais são unidades abstratas sujeitas a variações contextuais e a diferentes tipos de implementação (Cruz; Frota, 2010).[35]

Na língua portuguesa, os acentos tonais, sinalizados com um asterisco (*), são atribuídos, de forma geral, às sílabas tônicas (ou proeminentes), que caracterizam as palavras fonológicas (Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). As sílabas do português têm, como núcleo, a vogal, que é o elemento de maior sonoridade, e, como dominados, as consoantes e os glides que a rodeiam (Bisol, 2014). Há a possibilidade de os acentos tonais serem constituídos de um tom alto ou baixo (acentos monotonais ou simples), ou da combinação de dois tons (acentos bitonais ou complexos), em que apenas um deles é o principal. Os acentos tonais, no português, são ilustrados na Figura 6, de acordo com o sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015).[36]

[35] Embora os contornos entoacionais específicos de uma língua ou dialeto possam apresentar diferentes realizações fonéticas em outras línguas ou dialetos, em virtude de variações contextuais e de distintos tipos de implementação (Cruz; Frota, 2010), a língua, segundo o ponto de vista teórico adotado dentro do Gerativismo, não é concebida como um objeto social, mas como uma representação mental capaz de gerar as estruturas a partir de propriedades abstratas (Kato, 2005). Esse aspecto é pertinente à análise da entoação em ambientes computacionais, já que o desafio da síntese de fala não consiste somente na reprodução de especificidades acústicas, mas também na modelagem da lógica interna que organiza a estrutura dessa melodia.

[36] Figuras disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_tr_pa.html. Acesso em: 28 de maio de 2024.

Figura 6 – Acentos tonais do português

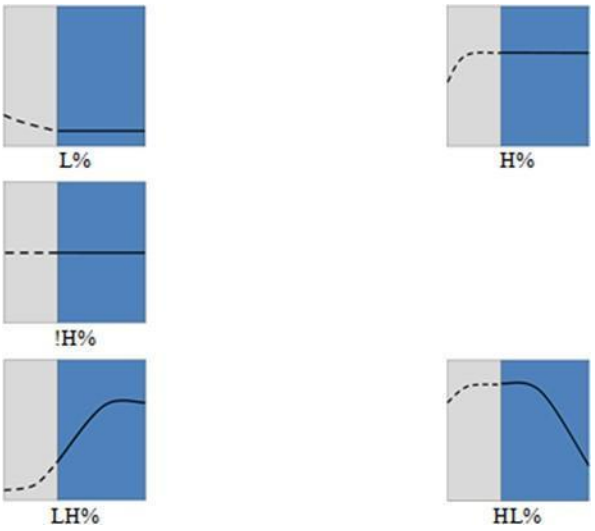


Fonte: Frota, Oliveira, P., Cruz e Vigário (2015).

Os tons associados a fronteiras prosódicas, por sua vez, classificam-se em acentos frasais (*phrase accents*) e tons de fronteira (*boundary tones*). Contudo, apenas os últimos são pertinentes à análise entoacional da declaração neutra no português brasileiro, visto que os acentos frasais, notados como H- e L-, ou somente H e L, não são vistos em fronteiras de enunciados declarativos neutros (Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008). Geralmente transcritos com o símbolo %, os tons de fronteira costumam ocorrer no término de uma frase entoacional (portadora de um contorno melódico particular), embora, em algumas línguas, possam surgir no início desse domínio prosódico. No português, eles tendem a ser atribuídos à direita de uma frase entoacional, como em várias outras línguas naturais. Essa atribuição é ilustrada na Figura 7, conforme o sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015).[37]

[37] Figuras disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_tr_nc.html. Acesso em: 22 de janeiro de 2025.

Figura 7 – Tons de fronteira do português



Fonte: Frota, Oliveira, P., Cruz e Vigário (2015).

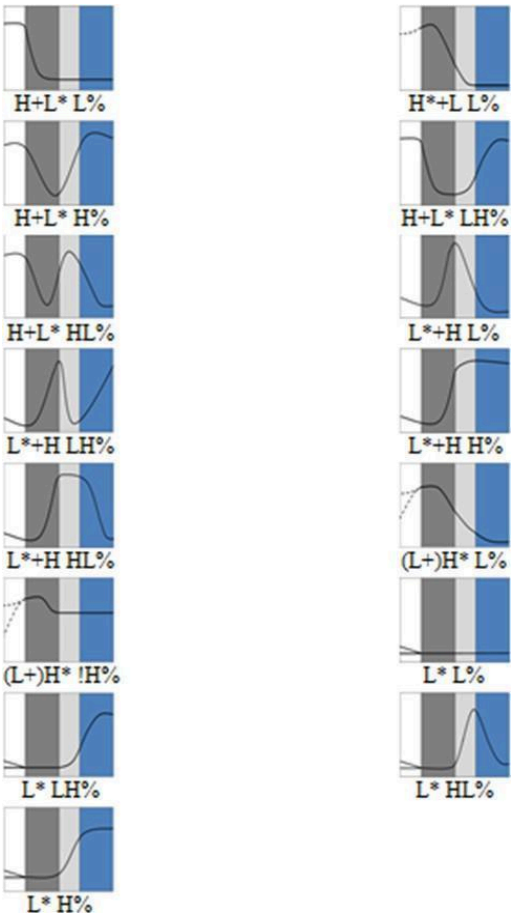
Segundo Frota, Oliveira, P., Cruz e Vigário (2015), as variedades do português utilizam as fronteiras tonais contrastivas de maneira mais restrita, quando observadas em comparação com outras línguas e variedades românicas. Os autores ainda apontam que as instâncias ocasionais de um tom inicial de fronteira de frase entoacional (%H) são descritas, por exemplo, no português falado em Lisboa, geralmente em uma distribuição complementar com outro tipo de pico inicial (como H*). No português brasileiro, por outro lado, um tom baixo pode, em alguns casos, marcar a fronteira direita de um constituinte de foco inicial (L-).[38]

Conforme exposto nos parágrafos precedentes, a estrutura tonal dos enunciados é caracterizada por uma combinação de acentos tonais e de tons de fronteira, de modo que o contorno melódico resultante é obtido por meio de interpolações fonéticas entre cada um desses pontos tonais (Collischonn, 2010). Além do conceito de acento tonal e de tom de fronteira, uma noção de suma relevância para o trabalho é a de contorno entoacional nuclear, que, em línguas como o português, constitui "a melodia na sílaba nuclear e na(s) sílaba(s) postônica(s) subsequente(s)" (Frota; Butler; Vigário, 2014, p. 199, tradução nossa).[39] No português, tanto a recursividade sintática quanto a proeminência mais forte nos sintagmas ocorrem à direita (Abaurre, 1996; Galves; Abaurre, 2002; Bisol, 2014). Nesse sentido, em uma relação entre a Sintaxe e a Fonologia, Frota, Oliveira, P., Cruz e Vigário (2015) apontam que a estrutura prosódica da língua portuguesa é caracterizada por frases prosódicas à direita,

[38] Informações disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_tr_bt.html. Acesso em: 26 de dezembro de 2024.
[39] *the melody on the nuclear syllable and subsequent post-tonic syllable(s)*.

com os eventos tonais relacionados a elas.[40] Sendo assim, o acento tonal nuclear tende a ser manifestado na fronteira direita da frase entoacional, exceto em frases com alguma focalização prosódica, que podem ter um núcleo inicial.[41] Os contornos entoacionais nucleares da língua portuguesa são ilustrados na Figura 8, conforme o sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015).[42]

Figura 8 – Contornos entoacionais nucleares do português



Fonte: Frota, Oliveira, P., Cruz e Vigário (2015).

A segunda perspectiva teórica em que a pesquisa se embasa, como anunciado no começo da subseção, é a Fonologia Prosódica, proposta por Nespor e Vogel (1986, 2007) e

[40] Com base em Frota (2000) e Fernandes (2007), o trabalho de Fernandes-Svartman (2024b) afirma que o contorno entoacional nuclear cumpre a função de destaque e assinala a principal proeminência frasal associada à palavra nuclear, correspondente à última palavra fonológica da frase entoacional em enunciados neutros do português, além de desempenhar um papel importante na diferenciação de tipos de enunciados, como em declarações e perguntas de sim/não, que podem ser segmentalmente idênticas, mas se diferenciarem por meio do contorno entoacional nuclear.

[41] Informações disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_tr_nc.html. Acesso em: 26 de dezembro de 2024.

[42] Figuras disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_tr_nc.html. Acesso em: 26 de dezembro de 2024.

também formulada no âmbito do Gerativismo. Essa abordagem proporciona às línguas regularidade e previsibilidade na organização sonora, por meio de regras fonológicas que se aplicam ou se bloqueiam dentro de certos constituintes ou em fronteiras específicas (Soncin; Tenani, 2016). Os princípios reguladores da hierarquia prosódica, aplicados às línguas naturais, são listados a seguir.

> i) cada unidade da hierarquia prosódica é composta de uma ou mais unidades da categoria imediatamente mais baixa;
> ii) cada unidade está exaustivamente contida na unidade imediatamente superior de que faz parte;
> iii) os constituintes são estruturas n-árias;
> iv) a relação de proeminência relativa, que se estabelece entre nós irmãos, é tal que a um só nó se atribui o valor forte (s) e a todos os demais o valor fraco (w) (Bisol, 2014, p. 260-261).

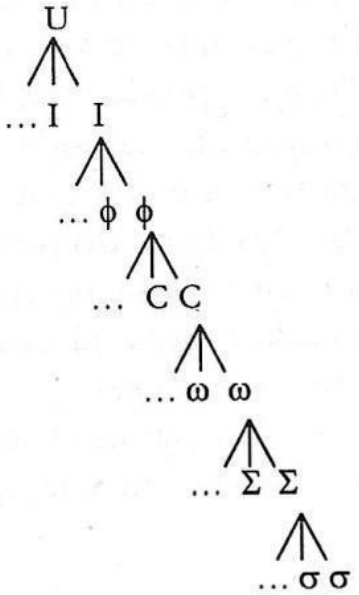
Os domínios (ou constituintes) incluídos na hierarquia prosódica do modelo de Nespor e Vogel (1986, 2007) são, em ordem crescente, o enunciado fonológico (U – *phonological utterance*), a frase entoacional (I – *intonational phrase*), a frase fonológica (ϕ – *phonological phrase*), o grupo clítico (C – *clitic group*), a palavra fonológica (ω – *phonological word*), o pé (Σ – *foot*) e a sílaba (σ – *syllable*). Eles não são necessariamente isomórficos aos outros constituintes gramaticais (Tenani, 2017).

A título de exemplo, a palavra fonológica não é sempre isomórfica à saída (*output*) do componente responsável pela formação de palavras (Abaurre; Wetzels, 1992). Com base na distinção entre a palavra morfológica e a palavra fonológica feita por Camara Jr. (1969, 1975), Bisol (2004) aponta que a palavra morfológica abrange tanto as palavras lexicais, como substantivos, adjetivos e verbos, que pertencem às classes abertas, quanto as palavras funcionais, como preposições, conjunções e determinativos, que pertencem às classes fechadas. Assim, em uma estrutura composta como "guarda-chuva", há uma palavra morfológica ("guarda-chuva") e duas palavras fonológicas dotadas de uma tonicidade primária ("guarda" e "chuva").[43]

Massini-Cagliari (1999) afirma que, segundo a teoria prosódica, as línguas têm regras específicas para a construção de constituintes, mas o consenso entre as línguas é que cada constituinte superior é formado a partir da combinação de unidades imediatamente inferiores. As relações entre os constituintes prosódicos são retratadas na Figura 9.

[43] De acordo com Vigário (2007, 2010), uma forma como "guarda-chuva" constitui um grupo de palavra prosódica (PWG – *prosodic word group*). Trata-se de um domínio prosódico discutido, em análises fonológicas subsequentes, por Toneli (2014) e Toneli, Abaurre e Vigário (2018).

Figura 9 – Representação das relações entre os constituintes prosódicos



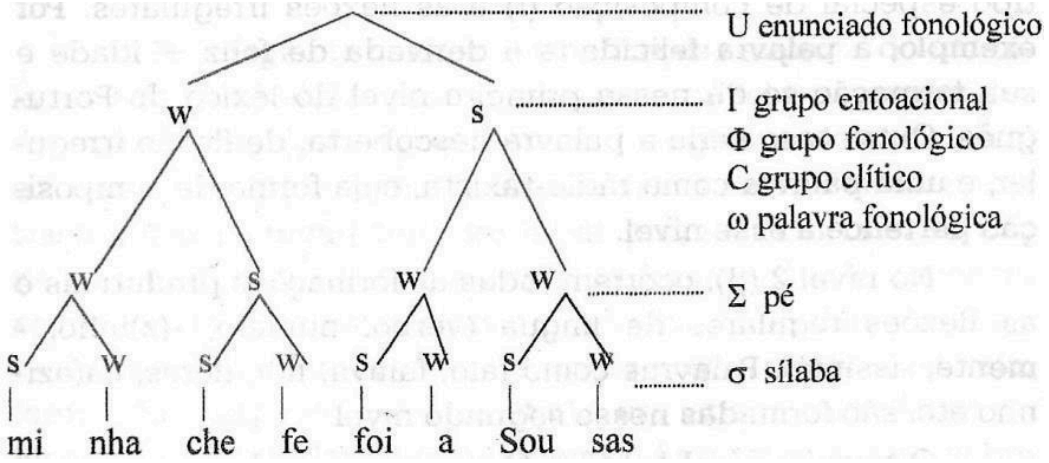
Fonte: Massini-Cagliari (1999, p. 127).

O modelo de Nespor e Vogel (1986, 2007) assume a ideia de que as "relações sintáticas são informações relevantes a partir das quais se configuram os constituintes prosódicos" (Tenani, 2017, p. 110). Em outros termos, os domínios prosódicos, sobretudo os hierarquicamente superiores, derivam da estrutura sintática (Abaurre; Wetzels, 1992), apesar de as estruturas prosódicas resultantes desse mapeamento poderem ser iguais ou diferentes das sintáticas (Tenani, 2002). Essa perspectiva é fundamental, por exemplo, para a definição do conceito de fraseamento prosódico, que se refere à "segmentação do contínuo de fala em unidades" (Serra, 2016, p. 48), e para o entendimento de que a interação entre a Sintaxe e a Fonologia é mediada pela estrutura prosódica (Massini-Cagliari, 1999).

Os domínios prosódicos podem ser representados por meio de uma estrutura de árvore ou de grade (Cagliari, 2002b). Nas Figuras 10 e 11, é apresentada, a partir da estrutura arbórea, uma representação dos exemplos "Minha chefe foi a Sousas" e "Pedro estuda na Universidade de Araraquara".[44]

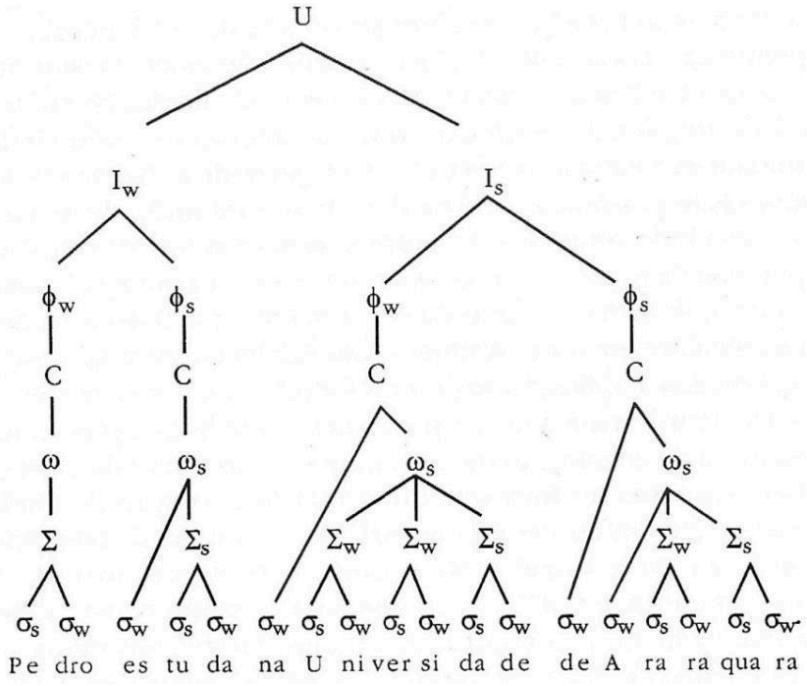
[44] Um exemplo como "Pedro estuda na Universidade de Araraquara" constitui, à primeira vista, uma frase entoacional e tem um significado completo. Entretanto, na representação de Massini-Cagliari (1999), "Pedro estuda" é uma frase entoacional, assim como "na Universidade de Araraquara", a fim de elucidar um caso de reestruturação de domínio prosódico. A frase entoacional básica "Pedro estuda na Universidade de Araraquara" é reestruturada em duas frases entoacionais mais curtas, por conta de diferentes motivações, cujas possíveis explicações se encontram em Nespor e Vogel (1986, 2007) e Frota (2000).

Figura 10 – Representação arbórea dos constituintes prosódicos do exemplo “Minha chefe foi a Sousas”



Fonte: Cagliari (2002b, p. 123).

Figura 11 – Representação arbórea dos constituintes prosódicos do exemplo “Pedro estuda na Universidade de Araraquara”



Fonte: Massini-Cagliari (1999, p. 128).

Os domínios prosódicos pertinentes à pesquisa incluem a palavra fonológica, a frase fonológica, a frase entoacional e o enunciado fonológico. Ao contrário da frase fonológica, da frase entoacional e do enunciado fonológico, o conceito de palavra fonológica já consta de uma apresentação prévia, que é retomada e expandida na próxima subseção, como parte da compreensão do domínio prosódico fundamental para a associação de tons no português brasileiro. Dessa forma, nas citações a seguir, são descritos apenas os algoritmos de formação

da frase fonológica, da frase entoacional e do enunciado fonológico. Em particular, os algoritmos de formação da frase fonológica e da frase entoacional, preliminarmente propostos por Nespor e Vogel (1986), são retirados da adaptação do estudo de Frota (2000) para o português europeu.

> Formação de Frase Fonológica (ϕ)
> a. **Domínio de ϕ**: um núcleo lexical X e todos os elementos em seu lado não recursivo, que ainda estão dentro da projeção máxima de X.
> b. **Reestruturação de ϕ**: inclusão opcional, obrigatória ou proibida de um ϕ ramificado ou não ramificado que é o primeiro complemento de X no ϕ que contém X (Frota, 2000, p. 56, grifos da autora, tradução nossa).[45]
>
> Formação de Frase Entoacional (*I*)
> a. **Domínio de I**: (i) todos os ϕs em uma sequência em que não esteja estruturalmente ligada à árvore da sentença (i.e., expressão em parênteses, perguntas-eco, vocativos, etc.); (ii) qualquer sequência restante de ϕs adjacentes em uma sentença raiz; (iii) o domínio de um contorno entoacional, cujas fronteiras coincidem com posições em que pausas gramaticais podem ser introduzidas em um enunciado.
> b. **Reestruturação de I**: (i) reestruturação de um *I* básico em *Is* mais curtos, ou (ii) reestruturação de *Is* básicos em um *I* maior. Fatores que desempenham um papel na reestruturação de *I*: comprimento dos constituintes, taxa de elocução e estilo interagem com restrições sintáticas e semânticas (Frota, 2000, p. 57, grifos da autora, tradução nossa).[46]
>
> *Formação de Enunciado Fonológico*
> I. *Domínio de U*
> O domínio de *U* consiste de todos os *Is*, correspondendo a X^n na árvore sintática.
> II. *Construção de U*
> Juntam-se, em uma ramificação n-ária de *U*, todos os *Is* incluídos em uma sequência delimitada pela definição do domínio de *U* (Nespor; Vogel, 2007, p. 222, grifos das autoras, tradução nossa).[47]

Com base nos princípios teóricos expostos por Ladd (1996), Callou e Serra (2012) destacam que, em diversas línguas, a frase entoacional constitui o principal domínio

---

[45] *Phonological Phrase (ϕ) Formation*
*a. **ϕ-domain**: a lexical head X and all elements on its non-recursive side, which are still within the maximal projection of X.*
*b. **ϕ-restructuring**: optional, obligatory, or prohibited inclusion of a branching or nonbranching ϕ which is the first complement of X into the ϕ that contains X.*

[46] *Intonational Phrase (I) Formation*
*a. **I-domain**: (i) all the ϕs in a string that is not structurally attached to the sentence tree (i.e. parenthetical expressions, tag questions, vocatives, etc); (ii) any remaining sequence of adjacent ϕs in a root sentence; (iii) the domain of an intonation contour, whose boundaries coincide with the positions in which grammar-related pauses may be introduced in an utterance.*
*b. **I-restructuring**: (i) restructuring of one basic I into shorter Is, or (ii) restructuring of basic Is into a larger I. Factors that play a role in I restructuring: length of the constituents, rate of speech, and style interact with syntactic and semantic restrictions.*

[47] *Phonological Utterance Formation*
. I. U domain
The domain of U consists of all the Is corresponding to X^n in the syntactic tree.
II. U construction
Join into an n-ary branching U all Is included in a string delimited by the definition of the domain of U.

prosódico em que ocorrem os alongamentos prévios às fronteiras, as inserções de pausas e os fenômenos entoacionais relevantes. Em outras palavras, segundo as autoras, esse constituinte prosódico é, por excelência, o domínio ao qual se vinculam os contornos entoacionais, por meio da distribuição de eventos tonais.

Uma questão que merece um tratamento detalhado é a validade das categorias analíticas tradicionais quando aplicadas a dados de fala sintética. A premissa é a de que as categorias prosódicas não são simplesmente o resultado de manifestações físicas humanas, mas a expressão de princípios de organização tonal que também podem ser observados em modelos treinados com um vasto conjunto de amostras linguísticas. A confirmação dessa hipótese reforça o estatuto fonológico das categorias melódicas e possibilita uma visão mais apropriada para os ambientes de interação entre os seres humanos e as máquinas.

Após a apresentação dos pressupostos teóricos do trabalho,[48] a subseção seguinte se dedica à entoação do português brasileiro, com foco no domínio prosódico de associação tonal e nos aspectos fonético-fonológicos de enunciados declarativos neutros. Para tanto, adota-se a visão integrada entre a Fonologia Entoacional Autossegmental e Métrica e a Fonologia Prosódica, inicialmente proposta em estudos como o de Frota (2000).

2.2 ENTOAÇÃO DECLARATIVA NEUTRA DO PORTUGUÊS BRASILEIRO: DOMÍNIO DE ASSOCIAÇÃO TONAL E ASPECTOS FONÉTICO-FONOLÓGICOS

Conforme mencionado na subseção anterior, os acentos tonais e os tons associados a fronteiras prosódicas são os eventos tonais básicos que caracterizam a estrutura fonológica da entoação de línguas como o português. Antes da exposição dos eventos tonais que costumam integrar a entoação declarativa neutra no português brasileiro, é imprescindível determinar o domínio prosódico de associação tonal dessa variedade linguística.

De acordo com a literatura prosódica acerca da temática (Massini, 1991; Massini-Cagliari, 1992, 1993; Fernandes, 2007b; Tenani; Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2010; Toneli, 2014; Fernandes-Svartman; Romano, 2017; Toneli; Abaurre; Vigário, 2018),[49] no português brasileiro, o domínio prosódico de atribuição de

[48] As investigações de Abaurre e Wetzels (1992), Massini-Cagliari (1999), Frota (2000), Cagliari (2002b), Tenani (2002, 2017), Bisol (2004), Vigário (2007, 2010), Callou e Serra (2012), Toneli (2014), Serra (2016), Soncin e Tenani (2016) e Toneli, Abaurre e Vigário (2018), referenciadas ao longo da subseção, examinam a língua portuguesa e aplicam, em algum momento, o modelo hierárquico de organização prosódica. A base teórica que orienta as referidas análises prosódicas é apresentada, de modo sistemático, por Nespor e Vogel (1986, 2007), cujos estudos definem os princípios estruturais da Fonologia Prosódica e estabelecem o enquadramento conceitual retomado em pesquisas posteriores sobre os aspectos acentuais, rítmicos e entoacionais do português.

[49] Apesar de os estudos de Massini (1991) e Massini-Cagliari (1992, 1993) não compartilharem a mesma

acentos tonais é a palavra fonológica, cuja propriedade elementar é a tonicidade (Vigário, 2003), manifestada pela marcação de acento na sílaba proeminente do constituinte fonológico em questão (Bisol, 2014). No nível fonético, o principal correlato acústico do acento é a duração (Massini, 1991; Massini-Cagliari, 1992, 1993), normalmente mensurada, no eixo temporal, em milissegundos (ms) e definida como o tempo utilizado na articulação de um domínio particular (Cristófaro Silva, 2015).

Em um trabalho anterior, Tenani (2002) determina que a frase fonológica é, no português brasileiro, o domínio prosódico de associação tonal. Já em um estudo posterior (Tenani; Fernandes-Svartman, 2008), a autora, em parceria com outra pesquisadora, considera a palavra fonológica como o domínio prosódico de atribuição de acentos tonais nessa variedade do português. Por outro lado, em alguns estudos contemporâneos (Vigário; Fernandes-Svartman, 2010; Toneli, 2014; Toneli; Abaurre; Vigário, 2018), é discutida a possibilidade de o grupo de palavra prosódica, proposto por Vigário (2007, 2010), constituir (ou não) o domínio de distribuição de acentos tonais na gramática do português brasileiro.

A pesquisa, em diálogo com as análises apresentadas em trabalhos anteriores, argumenta em favor da hipótese de que a palavra fonológica é o domínio mínimo de associação tonal no português brasileiro. Esse argumento é fundamentado na atribuição quase categórica de acentos tonais às palavras fonológicas dos enunciados declarativos neutros que compõem a amostra dos estudos citados, com uma validação experimental e estatística.

A palavra fonológica é determinada como o domínio prosódico de distribuição tonal, uma vez que, de acordo com os trabalhos referenciados acima, há uma tendência de os acentos tonais serem associados às palavras fonológicas em enunciados do português brasileiro, principalmente se elas constituírem a "cabeça" das frases fonológicas das quais fazem parte (Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2009), o que evidencia a alta densidade tonal dessa variedade linguística (Frota *et al.*, 2015; Frota; Moraes, 2016; Fernandes-Svartman, 2024a, 2024b). A densidade tonal corresponde à "proporção de acentos tonais em relação ao número de palavras prosódicas" (Fernandes-Svartman, 2012, p. 49). A atribuição de acentos tonais às palavras fonológicas "não cabeça" de frases fonológicas, por sua vez, é um fenômeno que ocorre de maneira opcional (Fernandes, 2007a, 2007b; Tenani;

---

perspectiva teórica adotada no estudo e não utilizarem a nomenclatura fonológica de acento tonal, eles oferecem uma análise sistemática e estatisticamente fundamentada da variação da F_0 em palavras fonológicas, com a articulação de questões rítmicas e acentuais também por meio de parâmetros físicos de duração e de intensidade. O rigor metodológico e o foco na mensuração acústica conferem a esses trabalhos uma referência relevante que corrobora a afirmação de a palavra fonológica ser o domínio prosódico de atribuição de acentos tonais no português brasileiro.

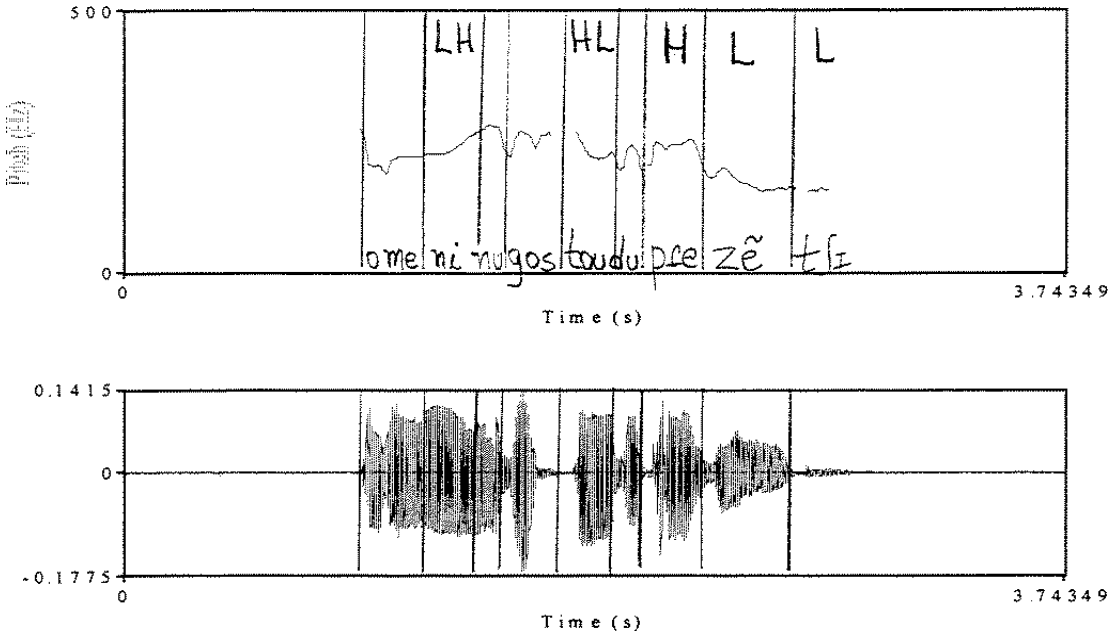
Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2009). Além disso, na variedade brasileira do português, costuma-se observar uma preferência pela alternância L H L H entre os tons (Tenani, 2002).

Tendo apontado o domínio mínimo de atribuição tonal na gramática do português brasileiro e as possibilidades de ocorrência de tons gramaticais nessa variedade linguística, os próximos parágrafos apresentam os principais eventos tonais atribuídos ao contorno entoacional de enunciados declarativos neutros. Para tanto, o trabalho de Tenani (2002) é tomado como a tese com a qual o estudo estabelece um diálogo preliminar, visto que as estruturas dos enunciados declarativos neutros da amostra são retiradas dessa pesquisa e ela é mencionada, no terceiro objetivo específico, como a principal investigação considerada para a comparação entre os aspectos da entoação declarativa neutra na fala sintética e na fala natural. De qualquer forma, um diálogo proveitoso com outras pesquisas sobre a estrutura entoacional do português brasileiro é estabelecido em momentos apropriados.

Tenani (2002) verifica que, em enunciados declarativos neutros do português brasileiro, o acento principal recai sobre a última sílaba tônica, à qual é associado o evento tonal H+L*, caracterizado por uma descendência na direção da curva entoacional, de acordo com as Figuras de 12 a 15. Segundo a autora, o tom baixo é alinhado à sílaba tônica e o tom alto à sílaba pretônica, independentemente de essa sílaba átona fazer parte ou não da mesma frase fonológica ou da mesma palavra à qual o tom baixo é atribuído, como exemplificam as Figuras 13 e 14. Além disso, a pesquisadora observa a presença de um tom L% associado à fronteira direita da frase entoacional, exceto quando a última sílaba acentuada ocupa a posição final desse domínio prosódico, o que resulta na ausência de material fônico necessário para a implementação do tom de fronteira, conforme ilustrado na Figura 15.[50] O contorno entoacional nuclear descendente também é identificado, para os enunciados declarativos neutros na variedade brasileira da língua portuguesa, em outros trabalhos prosódicos (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b).

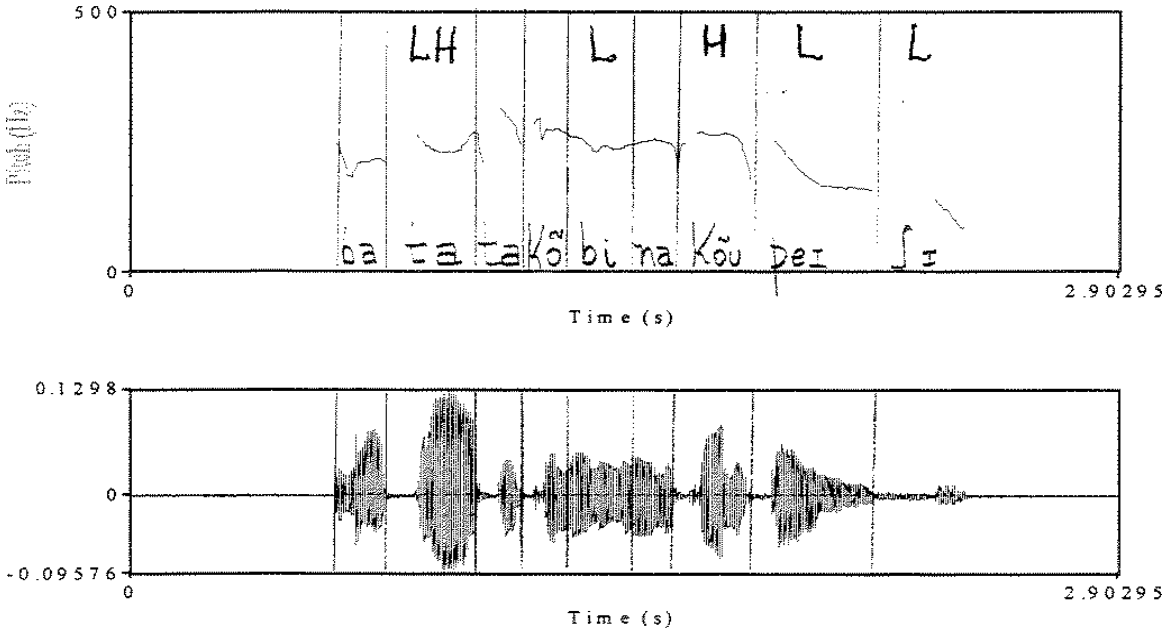
[50] Para a notação dos tons de fronteira, Tenani (2002) utiliza o símbolo "i", assim como Frota (2000), mas ressalta que esse segundo tipo de evento tonal também pode ser indicado pelo símbolo "%", como feito por Hayes e Lahiri (1991) e Ladd (1996). Em trabalhos posteriores (Tenani; Fernandes-Svartman, 2008; Soncin; Tenani, 2016; Soncin; Tenani; Berti, 2017, 2019; Tenani, 2017), a autora emprega o segundo símbolo para a notação de tons de fronteira.

Figura 12 – F_0 do enunciado declarativo neutro “O menino gostou do presente”



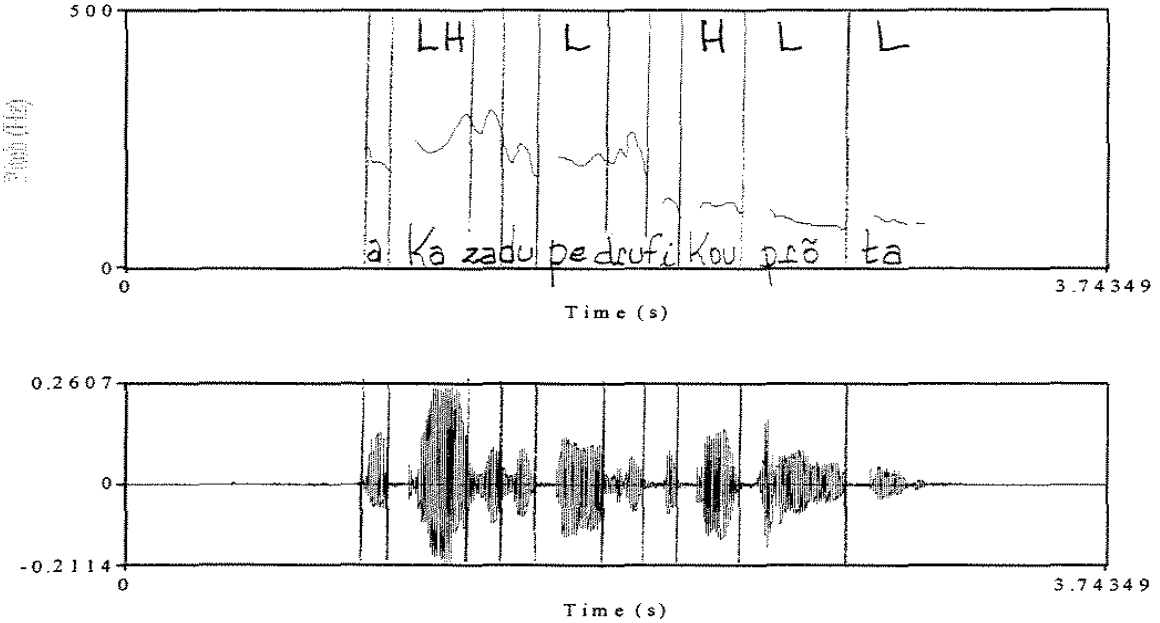
Fonte: Tenani (2002, p. 37).

Figura 13 – F_0 do enunciado declarativo neutro “Batata combina com peixe”



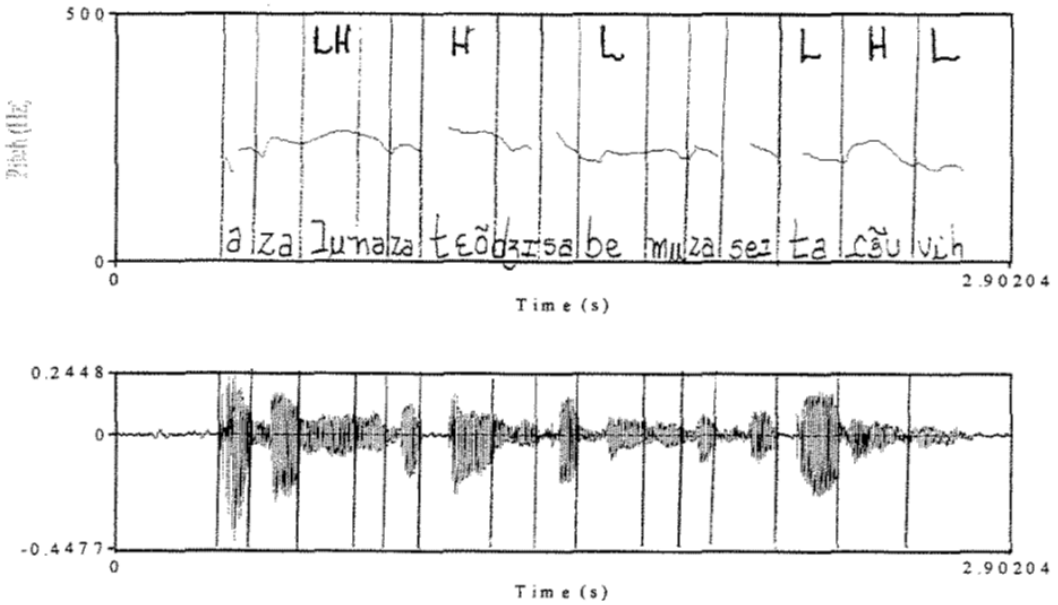
Fonte: Tenani (2002, p. 37).

Figura 14 – F_0 do enunciado declarativo neutro “A casa do Pedro ficou pronta”



Fonte: Tenani (2002, p. 38).

Figura 15 – F_0 do enunciado declarativo neutro “As alunas, até onde sabemos, aceitaram vir”



Fonte: Tenani (2002, p. 38).

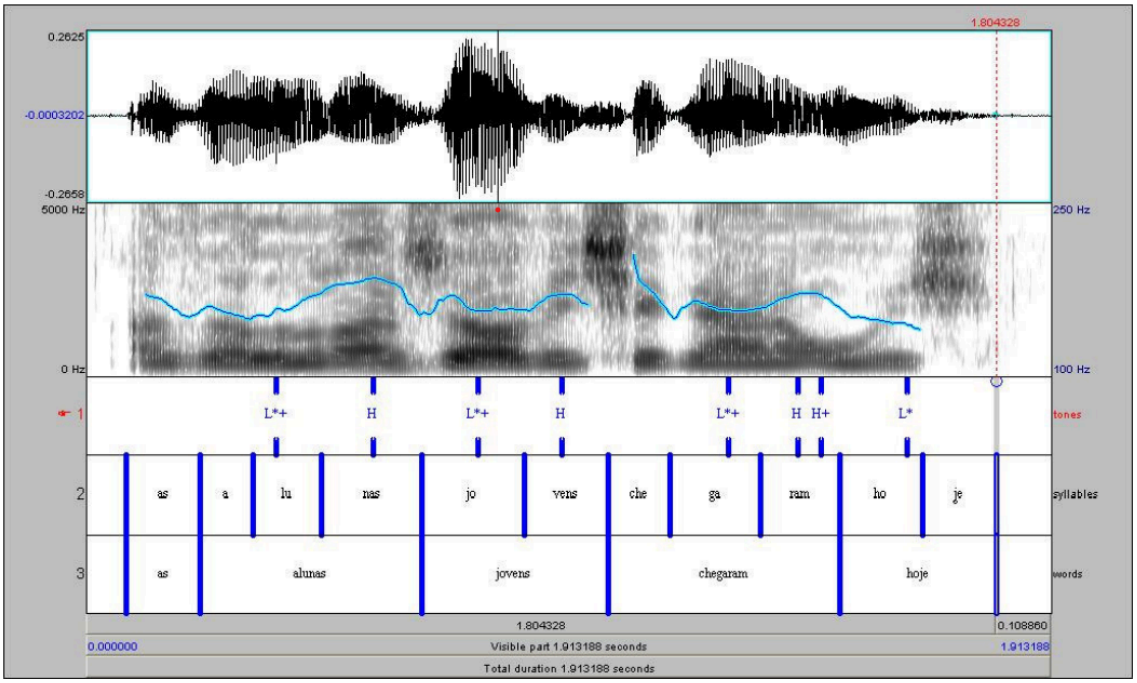
Castelo (2016) confirma a regularidade desse padrão nas variedades do português brasileiro e o descreve como uma melodia composta do acento nuclear H+L* e do tom de fronteira L%. O tom baixo se associa à sílaba nuclear da última palavra prosódica, o tom alto se alinha à sílaba pretônica imediatamente anterior à tônica e o tom de fronteira ocorre no término da frase entoacional. A autora observa, contudo, diferenças sutis de alinhamento e de escalonamento entre as regiões. Nas variedades da Paraíba e da Bahia, o movimento

descendente se inicia mais cedo e o tom baixo ocupa a maior parte da sílaba tônica, o que contrasta com o padrão do Centro-Sul, em que o movimento descendente se realiza integralmente na tônica. Nessas regiões, o contorno H+L* L% permanece estável e caracteriza a descida melódica final dos enunciados declarativos neutros. Fernandes-Svartman (2024a) confirma a generalização desse padrão em variedades faladas em João Pessoa (Paraíba), São Paulo (São Paulo) e Porto Alegre (Rio Grande do Sul), e observa que o contorno entoacional nuclear dos enunciados declarativos é sempre descendente, com a ocorrência de variação somente no nível fonético entre as regiões. Em um trabalho posterior, Fernandes-Svartman (2024b) reforça esses resultados e mostra que o padrão fonológico H+L* L% é uma propriedade comum do contorno entoacional nuclear dos enunciados declarativos neutros no português brasileiro, independentemente das diferenças regionais.

Em relação ao início da frase entoacional, que consiste em um aspecto relevante na análise da prosódia das línguas naturais por corresponder ao contorno entoacional pré-nuclear,[51] Tenani (2002) identifica um movimento ascendente na curva melódica. Em consonância com essa descrição, a literatura mostra que, embora não haja uniformidade quanto à notação fonológica empregada, a ascendência melódica inicial tende a ser um traço recorrente dos enunciados declarativos neutros (Cunha, 2000; Frota; Vigário, 2000; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Serra, 2009; Silvestre, 2012; Silvestre; Cunha, 2013; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Gomes da Silva *et al.*, 2016; Vieira, 2017; Fernandes-Svartman, 2024a). As Figuras 16 e 17 ilustram esse aspecto melódico, que evidencia a elevação inicial da curva da F_0, característica do contorno entoacional pré-nuclear. Nesse contexto, a notação fonológica L*+H descreve o padrão ascendente em que o tom baixo se associa à sílaba tônica, enquanto o tom alto se realiza na sílaba postônica, o que indica o alinhamento entre o acento tonal e a estrutura prosódica. Fernandes (2007a, 2007b) demonstra que esse tipo de acento tonal é amplamente produtivo no português brasileiro e recorrente no começo dos enunciados declarativos neutros, fato que reforça a adequação da notação para a representação fonológica do contorno inicial ascendente.

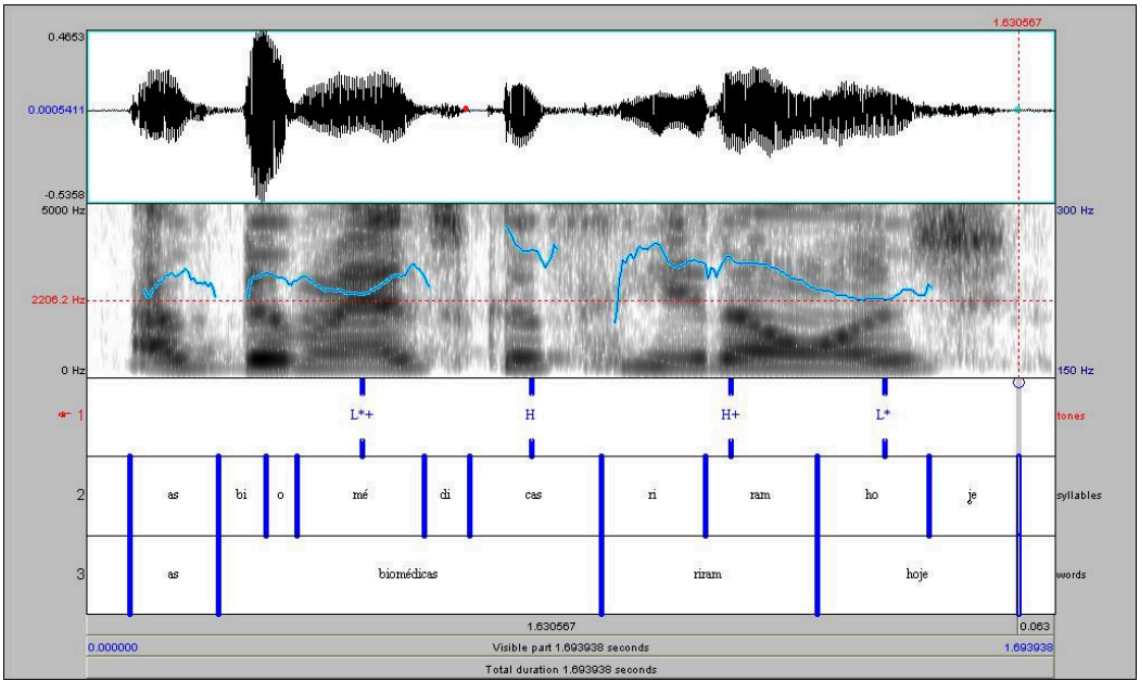
[51] Massini-Cagliari e Cagliari (2012) destacam que a programação prosódica inicial de um enunciado requer uma certa extensão temporal e deve considerar a entoação, o ritmo, a velocidade de fala e a acentuação. Os autores afirmam que os contornos melódicos da fala seguem determinados padrões. No caso das frases interrogativas, o final do contorno entoacional apresenta uma melodia ascendente e a organização da porção anterior do enunciado precisa levar em conta esse padrão, para que a entoação se desenvolva corretamente. Trata-se de um princípio vinculado ao processo neurolinguístico de produção da fala que também se aplica aos enunciados declarativos neutros.

Figura 16 – F_0 do enunciado declarativo neutro “As alunas jovens chegaram hoje”



Fonte: Fernandes (2007b, p. 197).

Figura 17 – F_0 do enunciado declarativo neutro “As biomédicas riram hoje”



Fonte: Fernandes (2007b, p. 197).

Castelo (2016) confirma a relevância desse padrão fonológico e revela uma distribuição regional sistemática dos acentos tonais pré-nucleares no português brasileiro. O estudo identifica dois tipos importantes de acentos iniciais: descendentes no Norte e ascendentes no Centro-Sul. Nas variedades centrais e meridionais, o contorno entoacional

pré-nuclear apresenta um movimento melódico inicial ascendente, caracterizado por um vale na sílaba tônica, seguido de uma elevação na sílaba postônica, comportamento típico do acento bitonal L*+H, amplamente preferido nessas regiões. No Norte, há uma maior variação, com a ocorrência de H+L* e de L*+H, o que indica um quadro de transição entre os padrões. Ainda que a autora aponte a possível influência de fatores extralinguísticos, como o nível de escolaridade, na adoção do acento tonal ascendente, o estudo ressalta que essa relação requer uma investigação mais detalhada. O fator regional, entretanto, mostra-se significativo e distingue duas áreas principais: o Norte, com uma alternância entre os acentos descendentes e ascendentes, e o Centro-Sul, com uma predominância estável dos acentos ascendentes. Essa distribuição confirma análises que descrevem a recorrência de movimentos melódicos ascendentes no começo dos enunciados declarativos neutros no português brasileiro (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Serra, 2009; Silvestre, 2012; Silvestre; Cunha, 2013; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Gomes da Silva *et al.*, 2016; Vieira, 2017; Fernandes-Svartman, 2024a), mesmo com notações fonológicas distintas. Além disso, Fernandes-Svartman (2024a) identifica, na porção pré-nuclear, o mesmo contorno entoacional ascendente em capitais como João Pessoa (Paraíba), São Paulo (São Paulo) e Porto Alegre (Rio Grande do Sul), fato que amplia a base empírica e comprova a extensão geográfica do padrão melódico L*+H.

No sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015), desenvolvido a partir da Fonologia Entoacional Autossegmental e Métrica e da Fonologia Prosódica, L*+H e H+L* L% são atestados no português brasileiro e representam, nessa ordem, os contornos entoacionais pré-nuclear e nuclear dos enunciados declarativos neutros. L*+H expressa o padrão de ascendência inicial, enquanto H+L* L% indica a configuração descendente final. Outros estudos, como os de Tenani (2002) e Moraes (2008), que examinam as variedades paulista e carioca, utilizam a representação L+H* para o contorno entoacional pré-nuclear, o que sugere uma tendência ascendente semelhante, ainda que com diferenças fonéticas de alinhamento tonal. Dentre os trabalhos que adotam explicitamente a notação fonológica L*+H para o português brasileiro, destacam-se, por exemplo, os de Fernandes (2007a, 2007b), Tenani e Fernandes-Svartman (2008), Frota *et al.* (2015) e Fernandes-Svartman e Romano (2017), cujos resultados convergem para a descrição de um padrão ascendente robusto no início dos enunciados declarativos neutros. A literatura indica que a ascendência melódica representada por L*+H é uma descrição fonológica consistente para o português brasileiro, já que ocorre em diferentes variedades dialetais (Fernandes-Svartman, 2024a; Castelo, 2016).

A produção bibliográfica concernente à prosódia do português brasileiro é extensa e madura em diferentes aspectos. No entanto, a interseção entre essa tradição e a investigação sistemática de saídas acústicas de síntese de fala baseada em IA permanece pouco explorada. Em outras palavras, embora haja contribuições significativas sobre os contornos entoacionais na fala natural, há uma necessidade de estudos que utilizem esses marcos teóricos para a avaliação e a teorização da entoação produzida por modelos modernos de síntese de fala. Diante da lacuna mencionada, há uma oportunidade teórica de testar a aplicação, sem qualquer perda analítica, das categorias e dos mecanismos construídos pelos falantes humanos em recursos computacionais.

A próxima subseção delineia os principais fundamentos da Linguística Computacional que embasam a pesquisa, bem como apresenta um panorama da síntese de fala, com ênfase nos aspectos históricos, nos principais métodos e nas abordagens de produção entoacional.

2.3 SÍNTESE DE FALA: UMA INTERFACE ENTRE A LINGUÍSTICA E A CIÊNCIA DA COMPUTAÇÃO

Além da Fonética e da Fonologia, utilizadas na pesquisa para a análise da entoação, a outra disciplina incluída no escopo do trabalho é a Linguística Computacional, que surge da interação entre os estudos linguísticos, computacionais e cognitivos (Ferreira; Lopes, 2017). Na investigação, a Linguística Computacional estabelece um diálogo com o Processamento de Linguagem Natural (PLN), que constitui um conjunto de métodos que tornam a linguagem humana acessível aos computadores, além de se integrar ao cotidiano por meio de recursos como a tradução automática, os filtros de *spam* em *e-mails*, os mecanismos de busca linguisticamente sofisticados e os sistemas de diálogo interativos (Eisenstein, 2019).

Conforme afirmam Ferreira e Lopes (2017), a Linguística Computacional, como uma parte da Linguística, tem se tornado cada vez mais relevante nas atividades de descrição e de análise linguística, pois permite testar as hipóteses por meio das informações sobre as línguas naturais contidas em bancos de dados empíricos. Por sua vez, como um subcampo da Ciência da Computação, a Linguística Computacional faz parte da pesquisa em IA e se concentra na compreensão e na produção de línguas humanas, com o objetivo de permitir a comunicação mediada por computador, a partir de recursos como os corretores ortográficos e gramaticais, os sistemas de reconhecimento e de síntese de fala e os tradutores automáticos. Sob a perspectiva das Ciências Cognitivas, os estudos em Linguística Computacional se dedicam a aspectos relacionados à interação entre a linguagem e o pensamento, assim como à

formulação de modelos do processamento linguístico no cérebro.

Para o desenvolvimento do trabalho, um conceito crucial é o de algoritmo, amplamente utilizado no domínio da programação. Como explicam Ferreira e Lopes (2021), um algoritmo é definido como um conjunto de instruções organizadas para resolver um problema, caracterizado por um número limitado de etapas e uma sequência bem definida. Em linhas gerais, os autores apontam que o processo de computação envolve o fornecimento de dados de entrada, que são processados por um algoritmo. Ao término da execução, o algoritmo gera os dados transformados.

Silva, A., Silva, F. e Chacon (2024) apontam que, na elaboração de qualquer IA, uma das etapas iniciais é a criação de um *corpus* para o treinamento do algoritmo, a partir do qual ela "aprende" os padrões, generaliza as informações e identifica as regularidades em novos conjuntos de dados fornecidos ao sistema. Segundo os autores, a criação de modelos computacionais que possibilitem a compreensão do processamento da linguagem humana pode gerar várias contribuições, pois permite a investigação das hipóteses envolvidas nesse processamento em diferentes níveis, o que facilita a proposição de soluções para diversos problemas.

Conforme Freitas (2022), o Aprendizado de Máquina (*Machine Learning*) utiliza os dados e as experiências anteriores para prever a melhor maneira de agir no futuro. Para a criação de um algoritmo eficiente, a autora destaca a necessidade do aprendizado com uma quantidade de dados diversos e representativos do que se deseja aprender. Os dados são essenciais para "alimentar" as máquinas, tanto em quantidade quanto em qualidade.

De acordo com Freitas (2022), nos últimos anos, o Aprendizado Profundo (*Deep Learning*), um tipo específico de Aprendizado de Máquina, tem apresentado avanços significativos no campo da IA, sobretudo no que se refere ao PLN. Esse tipo de aprendizado utiliza as redes neurais artificiais como a base da operação (Freitas, 2022), as quais tentam se assemelhar ao funcionamento do cérebro humano e consistem em "sistemas de processamento de informação formados pela interconexão de unidades simples de processamento, denominadas neurônios artificiais" (Von Zuben, 2003, p. 66).

Silva, A. (2020b) destaca a convergência entre a Computação e a Linguística em dois aspectos principais: a arquitetura e o processamento da linguagem, por um lado, e a análise de dados, por outro. Ambas as áreas compartilham a ideia de que a linguagem, seja humana ou artificial, é formada por blocos ou módulos inter-relacionados que atuam em conjunto para gerar um resultado. No caso da linguagem humana, esses módulos envolvem a produção de sentido, a formação de sentenças e de palavras e a associação de sons, ou seja, uma estrutura

que se assemelha ao funcionamento das linguagens de programação, em que os blocos de código interagem para construir algoritmos capazes de resolver tarefas específicas. Ademais, tanto na Linguística quanto na Computação, a análise de dados abrange a coleta, a organização e a modelagem de vastos conjuntos de informações, o que possibilita a identificação de padrões e a explicação de variações. O diálogo entre essas áreas contribui para a otimização da análise, da documentação e do armazenamento de dados linguísticos, além de favorecer a melhoria do processamento e da interpretação de línguas naturais. Essa interação, porém, não se restringe à organização e ao tratamento de dados, mas também impulsiona o desenvolvimento de ferramentas que utilizam a linguagem natural para diversas finalidades.

Recentemente, alguns estudos têm discutido a compatibilidade entre a IA e a Linguística Gerativa, como é o caso do trabalho de Portelance e Jasbi (2025). Os autores argumentam que essas abordagens compartilham fundamentos teóricos e objetivos científicos, uma vez que os modelos de linguagem neural podem ser compreendidos como sistemas formais de geração de linguagem que retomam os princípios da teoria formal da linguagem desenvolvida pela tradição gerativa. Além disso, os pesquisadores sustentam que tais modelos contribuem para a formulação de procedimentos de descoberta e para o avanço das investigações sobre a estrutura e o funcionamento da linguagem.

No campo das interfaces entre a Linguística e a IA, destacam-se a Fonética e a Fonologia na formulação de tecnologias voltadas ao PLN. Conforme observado por Othero (2006), essas disciplinas contribuem para o desenvolvimento de sistemas de reconhecimento e de síntese de fala, além de constituírem a base teórica de mecanismos de diálogo na modalidade falada da língua. Os aplicativos resultantes dessa integração são caracterizados por uma extensa gama de funcionalidades, que abrangem desde o reconhecimento de comandos vocais em dispositivos eletrônicos até a conversão automática de ditados orais em texto escrito. Os sistemas de síntese de fala, por sua vez, viabilizam a produção de voz a partir de dados textuais e a leitura oral de documentos digitais. Dessa forma, há o favorecimento da inclusão de pessoas com deficiência visual e a melhoria da acessibilidade tecnológica. Tais avanços evidenciam o papel da Linguística na elaboração de recursos computacionais que ampliam as formas de interação entre o ser humano e a máquina por meio da linguagem natural.

Em uma perspectiva complementar, Cristófaro Silva (2011) argumenta que a Fonologia oferece fundamentos indispensáveis ao aprimoramento das tecnologias de comunicação oral mediadas por computador. A autora sustenta que a elaboração de

instrumentos capazes de interagir com as pessoas pela fala exige um entendimento rigoroso da estrutura e da organização da sonoridade linguística. Essa compreensão deve integrar uma abordagem interdisciplinar, com a colaboração de especialistas de diferentes áreas na elaboração de modelos e de bases de dados adequados aos objetivos teóricos e tecnológicos. A pesquisadora também observa que o diálogo entre a Linguística e a Computação favorece o aprofundamento das reflexões sobre os princípios estruturais da linguagem e o desenvolvimento de ferramentas eficazes para o processamento sonoro em sistemas computacionais relacionados à comunicação.

Após delinear os principais fundamentos da Linguística Computacional que embasam a pesquisa e situar a investigação de síntese de fala no quadro da interface entre a Linguística e a Ciência da Computação, as próximas páginas são dedicadas a uma apresentação da tecnologia de síntese de fala. Exibe-se uma visão geral dessa tecnologia, com destaque aos aspectos históricos, aos principais métodos e às abordagens de produção entoacional.

A pesquisa adota o conceito segundo o qual a síntese de fala é a "produção de fala por máquinas, por meio da fonetização automática das frases a serem pronunciadas" (Dutoit, 1997, p. 13, grifos no original, tradução nossa).[52] Embora possa ser vista como uma tecnologia contemporânea, é importante informar ao leitor que as tentativas de reprodução vocal são identificadas em diferentes momentos históricos.

O trabalho reconhece que, desde os séculos anteriores, há alguns registros de iniciativas que buscam reproduzir a fala humana por meio de máquinas, como demonstra o trabalho de Barbosa (2001) e Barros (2002). No entanto, o objetivo da subseção é fornecer ao leitor um panorama da síntese de fala, com foco em algumas das principais "máquinas falantes" criadas ao longo da história, mesmo com as limitações impostas pela engenharia da época, a fim de permitir que ele reflita sobre as transformações técnico-científicas ocorridas na sociedade quanto à tecnologia de fala estudada na pesquisa.

Historicamente, a primeira tentativa de construção de um autômato falante remonta ao século XVIII, com a máquina do barão húngaro Wolfgang von Kempelen (Kempelen, 1791), visível na Figura 18,[53] construída entre 1769 e 1791 e motivada pela preocupação com a educação inclusiva (Barbosa, 2022). Ela desempenha o papel de simular a articulação, ou seja, o último estágio do mecanismo de produção da fala, por meio da ação manual de suas estruturas (Barbosa, 2022).

---

[52] *production of speech by machines, by way of the automatic phonetization of the sentences to utter*.
[53] Os créditos da fotografia são do Deutsches Museum (Hans-Joachim Becker).

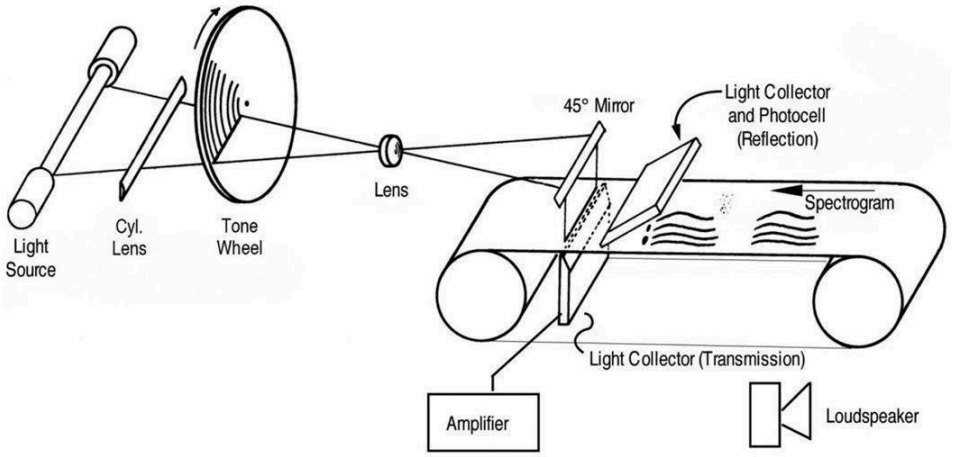
Figura 18 – Máquina de falar “Kempelen”

Fonte: Reprodução de Leibniz Association (20--).

A máquina de von Kempelen se constitui de um fole, de um bocal cujo volume varia com o movimento da mão esquerda para a produção dos segmentos vocálicos, de narinas e de apitos acionados por alavancas controladas pela mão direita para a produção dos segmentos consonantais (Barbosa, 1999). Trata-se da verdadeira tentativa de compreensão e de reprodução dos sons da linguagem articulada (Barbosa, 1999).

O período da Segunda Guerra Mundial (1939-1945) se destaca como significativo para o avanço das iniciativas de criação de um sistema de síntese de fala. Durante o contexto histórico mencionado, há o surgimento do espectrógrafo, que, além de intensificar a pesquisa sobre a percepção de fala (Berti, 2008), permite, no ano de 1951, a construção do Pattern Playback (Barbosa, 2022), exibido na Figura 19.

Figura 19 – Representação do Pattern Playback



Fonte: Rubin e Goldstein (20--).

O Pattern Playback, criado por Cooper *et al.* (1951), representa um notável progresso na área de fala sintética, em um contexto histórico marcado por conflitos políticos e sociais. Uma breve descrição do funcionamento do Pattern Playback, escrita por Barbosa (2022) a partir da leitura do estudo realizado pelos desenvolvedores da máquina, é apresentada abaixo.

> Através de um sistema formado por um disco, um jogo de lentes e um espelho a 45 graus, é possível jogar luz num espectrograma invertido em termos de preto e branco e gerar níveis de corrente elétrica proporcionais ao padrão de cinza pela transdução realizada por uma célula fotoelétrica, permitindo a reprodução do som. É fácil imaginar que a manipulação sintética dos traçados tenha permitido obter os primeiros trechos de fala sintética. Seguiram-se décadas de experimentos que auxiliaram no desenvolvimento das teorias de percepção da fala, incluindo a relevância relativa de parâmetros acústicos como formantes e transições formânticas, que são pistas relacionadas à produção das consoantes, especialmente as oclusivas (Barbosa, 2022, p. 80).

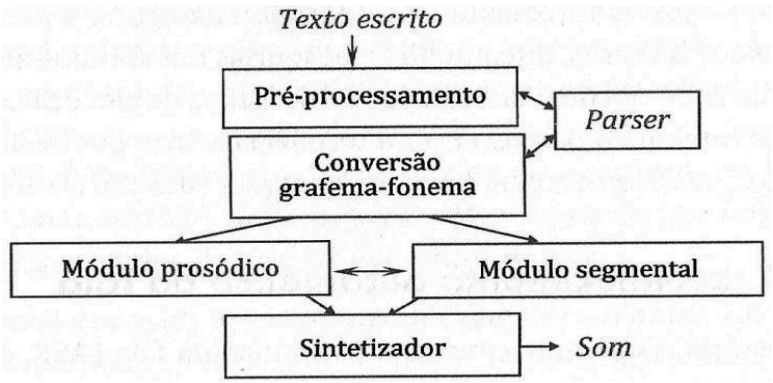
Na contemporaneidade, para que a síntese de fala seja efetuada, Barbosa (2022) aponta a necessidade de reunir e de organizar, de acordo com critérios linguísticos, tecnológicos e de produção de fala, vários elementos, blocos e subestruturas linguísticas e sonoras analisadas previamente. De acordo com o autor, a síntese de fala realizada com base no texto escrito consiste no modelo mais empregado nos dias de hoje e incorpora regras linguísticas. Esse modelo de síntese de fala, cuja tarefa simula aquela que o ser humano realiza ao ler um texto escrito em voz alta (Madureira, 1999), favorece a existência de duas linhas de pesquisa, conforme explicado abaixo.

> A possibilidade de realizar a síntese da fala a partir do texto que, como o próprio nome sugere, significa emitir os sons da fala a partir de uma representação textual da mensagem, desde cedo suscitou duas linhas de pesquisa. A primeira linha busca reproduzir da melhor forma possível um sinal acústico que *pareça* com o sinal da fala (chamaremos de abordagem "fazer-parecido"). A segunda linha procura obter sinal acústico a partir das causas que o propiciaram, reproduzindo o mecanismo fonatório da forma *como* ele funciona no ser humano (chamaremos de abordagem "fazer-como-se-fosse") (Barbosa, 1999, p. 25, grifos do autor).

A partir das explicações conceituais de Ferreira e Lopes (2017, 2021) e Freitas (2022), mencionadas em páginas precedentes, pode-se aplicar à tecnologia de síntese de fala a ideia de que um texto escrito é processado por uma sequência de instruções algorítmicas que o transformam em fala audível. O sistema de computador deve ser "alimentado" com dados linguísticos de alta qualidade baseados na fala de somente um sujeito, cuja escolha se pauta em uma avaliação da qualidade de voz, da precisão articulatória, da habilidade de variação da taxa de locução e da capacidade de expressão clara e natural do falante em uma situação de leitura (Madureira, 1999). Esse procedimento, aliado a uma sólida análise linguística (Madureira, 1999), é necessário para garantir uma conversão bem-sucedida do texto escrito em fala audível.

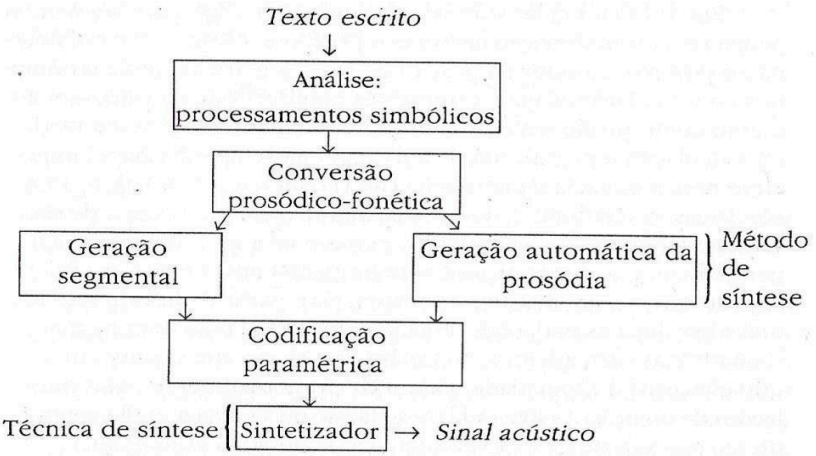
Segundo Barbosa (2022), em um modelo de síntese de fala, o sintetizador, que é a última etapa do processo de conversão texto-fala, tem a função de coletar as informações integradas do módulo segmental e do módulo prosódico e de encaminhá-las a uma placa sonora responsável pela reprodução do sinal de fala. As Figuras 20 e 21 apresentam um diagrama e um esquema de um sistema de síntese de fala a partir de texto escrito, extraídos de trabalhos do autor.

Figura 20 – Diagrama geral de um sistema de síntese de fala a partir do texto escrito



Fonte: Barbosa (2022, p. 81).

Figura 21 – Esquema geral de um sistema de síntese de fala

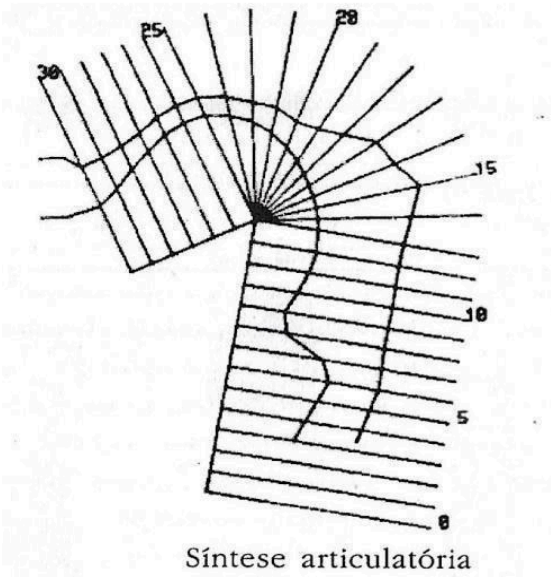


Fonte: Barbosa (1999, p. 27).

A depender da natureza das informações integradas pelos módulos segmental e prosódico, há métodos utilizados para a realização da síntese de fala (Barbosa, 2022). Esses métodos surgem e se consolidam ao longo do tempo, mas apresentam disponibilidade, custo e complexidade distintos, assim como aplicações específicas. Cada método dispõe de particularidades, que são explicadas a seguir.

A síntese articulatória, ilustrada na Figura 22, modela, em termos matemáticos, o trato vocal e simula os movimentos dos articuladores para se obter a saída acústica (Albano *et al.*, 1999; Barbosa, 2022). Apesar do potencial de alta qualidade, o elevado custo computacional e a complexidade de implementação restringem a aplicação dessa técnica principalmente a contextos acadêmicos (Albano *et al.*, 1999). Esse método possibilita a modelagem detalhada do trato vocal a partir da posição dos articuladores e das áreas de constrição no plano sagital (Albano *et al.*, 1999; Barbosa, 2022). No entanto, por causa da dificuldade de se observar diretamente a articulação, a síntese articulatória não se mostra uma opção prática para aplicações comerciais de conversão texto-fala (Albano *et al.*, 1999).

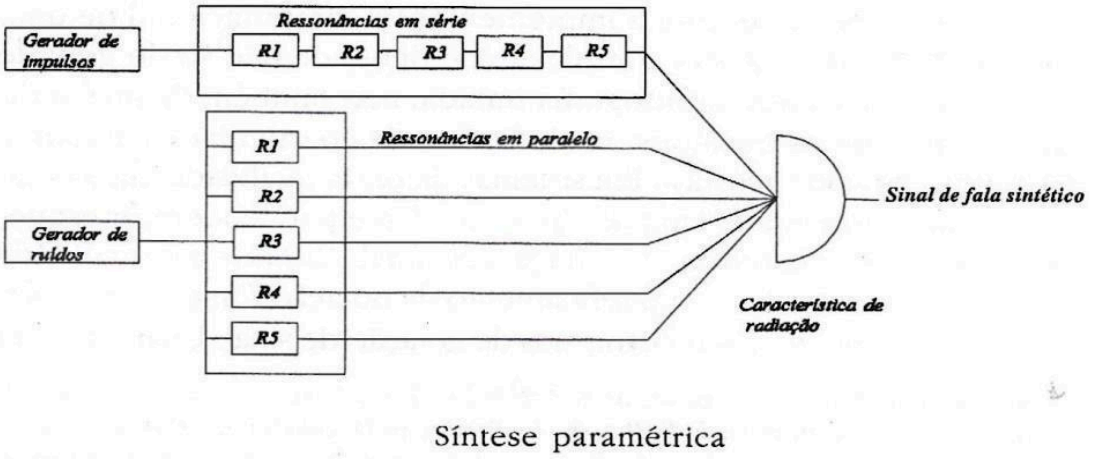
Figura 22 – Representação da simulação de fala na síntese articulatória



Fonte: Albano *et al.* (1999, p. 86).

Já a síntese paramétrica gera a fala de acordo com parâmetros referentes à fonte sonora e ao trato vocal (Barbosa, 2022). Esse método, exemplificado na Figura 23, utiliza um processo eletrônico para simular os parâmetros acústicos que descrevem o sinal de fala (Albano *et al.*, 1999). A técnica mais comum é a síntese de formantes, em que a saída de um gerador de pulsos (ou de ruídos), representante da fonte sonora da fala, atravessa uma série de filtros correspondentes às ressonâncias do trato vocal (Albano *et al.*, 1999).

Figura 23 – Representação da simulação de fala na síntese paramétrica

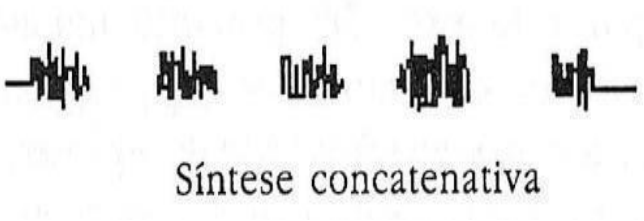


Fonte: Albano *et al.* (1999, p. 86).

Por outro lado, a síntese concatenativa, representada na Figura 24, populariza-se em virtude da relação custo-benefício e da escalabilidade, além de combinar um conjunto de

fragmentos de fala (difones, trifones e polifones) pré-gravados, de forma a preservar as transições sonoras (Albano *et al.*, 1999; Barbosa, 2022). No passado, pouca atenção é dedicada à prosódia e as unidades apresentam características prosódicas fixas, o que compromete a naturalidade (Aquino, 1997). As limitações de processamento e de memória dos computadores condicionam o tamanho e o número de unidades nos bancos de dados e reduzem o detalhamento fonético possível (Albano *et al.*, 1999). Atualmente, unidades maiores, que podem incluir enunciados inteiros, também integram a síntese concatenativa (Barbosa, 2022).

Figura 24 – Representação da simulação de fala na síntese concatenativa



Fonte: Albano *et al.* (1999, p. 86).

O sintetizador se integra a outras etapas essenciais para a conversão de texto escrito em fala audível, como o pré-processamento, a conversão grafema-fonema, o *parser* e o módulo prosódico (Albano *et al.*, 1999; Madureira, 1999; Barbosa, 2022). A etapa de pré-processamento transforma o texto em uma representação fonológica e possibilita a pronúncia correta de abreviações, de números e de símbolos gráficos. O módulo de conversão grafema-fonema realiza a transposição de texto para fonemas e utiliza um conjunto de regras e de procedimentos para o tratamento de exceções.[54] O *parser* realiza a análise morfossintática, resolve as ambiguidades e identifica os padrões que não se enquadram nas regras gerais. Por fim, o módulo prosódico define o ritmo e a entoação e ajusta a sequência de articuladores ou os trechos de som, de tal modo a assegurar a pronúncia adequada do texto (Barbosa, 2022).

Conforme Casanova *et al.* (2024), em razão dos avanços expressivos no domínio do Aprendizado Profundo, os sistemas de síntese de fala se destacam nos últimos anos (Casanova, 2019, 2022; Casanova; Shulby; Aluísio, 2021; Casanova *et al.*, 2021, 2022). Com a IA, a síntese neural obtém os padrões acústicos a partir de *corpora* e supera as restrições históricas, financeiras e computacionais. Nesse contexto, o estudo *WaveNet: A Generative*

[54] Abaurre (1999, 2019) destaca que os dados escritos nem sempre são uma evidência direta da pronúncia. No entanto, a escrita alfabética preserva uma certa correspondência entre os fonemas e os grafemas, de modo que, em alguns casos, as características da fala podem se refletir indiretamente na forma escrita.

Model for Raw Audio, elaborado por van den Oord *et al.* (2016), apresenta o WaveNet, um modelo voltado à produção de formas de onda de áudio em um estado bruto. Segundo os autores, o modelo demonstra uma elevada capacidade de reprodução de vozes sintéticas com uma naturalidade superior à dos principais métodos de síntese de fala. A pesquisa indica, ainda, que o sistema reduz a discrepância subjetiva em relação à fala humana em mais da metade dos casos. A rede neural também realiza a síntese de outros tipos de sinais sonoros, como músicas, e apresenta exemplos expressivos de peças de piano geradas de forma automática. Desenvolvido pela DeepMind, o WaveNet viabiliza novas aplicações em síntese de fala, geração musical e modelagem de áudio.

A pesquisa investiga o módulo prosódico de um sistema de síntese de fala, especialmente a geração da entoação, uma vez que, na literatura nacional, por exemplo, há uma carência de estudos que abordem a interseção entre a prosódia e a síntese de fala. Em uma revisão integrativa de pesquisas brasileiras sobre a relação entre esses domínios, distribuídas nas áreas de Engenharia Elétrica, de Engenharia de Computação, de Engenharia de Teleinformática, de Engenharia Mecatrônica, de Engenharia Eletrônica e de Computação, de Ciência da Computação e de Linguística, Galdino e Oliveira Jr. (2023) evidenciam que:

> [...] a prosódia tem sido considerada como essencial para o desenvolvimento da síntese de fala, a partir de informações linguísticas aliadas ao campo das áreas das engenharias. A presença da Linguística nesses trabalhos demonstra que ela é uma área importante, uma vez que contribui para uma voz sintética mais expressiva e aceita pelos usuários, especificamente nos níveis prosódicos (Galdino; Oliveira Jr., 2023, p. 10).

Com relação à entoação, que, juntamente com o tom e a tessitura, faz parte dos elementos prosódicos da melodia da fala (Cagliari, 1992), Madureira (1999) observa que duas perspectivas, de orientação oposta, têm sido implementadas em sistemas voltados à síntese de fala: a foneticamente orientada e a fonologicamente orientada. Essas perspectivas, cujas diferenças não envolvem aspectos como a expressividade, a sistematicidade, a arbitrariedade ou a especificidade dos contornos entoacionais (Madureira, 1999), são discutidas a seguir.

De acordo com Madureira (1999), a abordagem de base fonética pressupõe um contorno entoacional específico para cada nível linguístico e lida com a ideia de sobreposição de relevos em domínios linguísticos hierarquicamente especificados (frase, sentença, sintagma, palavra e sílaba) para gerar os contornos da F_0 que variam continuamente e têm um formato global. Os modelos fonéticos, cuja natureza é indutiva, trabalham com a gradiência e relacionam o gradiente e o discreto de modo quantitativo. O módulo entoacional em sistemas

de síntese de fala orientados por tais modelos especifica os contornos da F_0 quantitativamente. Esses modelos anulam a distinção entre a Fonética e a Fonologia, em favor de uma Fonologia Dinâmica operante no contínuo, o que elimina a representação no nível simbólico.

Já a abordagem de base fonológica, segundo Madureira (1999), trabalha com a noção de sequência. Os contornos entoacionais consistem em sequências de elementos fonológicos discretos, chamados de tons. Os modelos fonológicos, cuja natureza é dedutiva, lidam com as categorias discretas e relacionam o discreto e o gradiente de modo qualitativo. O módulo entoacional em sistemas de síntese de fala orientados por tais modelos especifica os contornos da F_0 de maneira linear, como sequências de tons categoricamente distintos e concatenados entre si. Em geral, há dois estágios nesses modelos: o de conversão do texto escrito em Fonologia e o de conversão da Fonologia em Fonética, em que os tons são mapeados em contínuos da F_0.

Conforme informado ao leitor, o Google Cloud Text-to-Speech, oferecido pelo Google Cloud Platform e desenvolvido com base na experiência em síntese de fala da DeepMind,[55] é o sistema utilizado na pesquisa, que afirma oferecer aos usuários benefícios como uma fala de alta fidelidade, uma seleção de voz mais ampla e uma voz exclusiva.[56] Embora o objetivo do estudo não seja a criação de um novo sistema de síntese de fala, o trabalho manifesta uma concordância com a posição de Madureira (1999), que, ao discutir os modelos fonéticos e fonológicos, defende a presença de um componente fonológico e de um componente de implementação fonética dinâmica. Nesse sentido, a autora considera plausível a manutenção de um nível fônico simbólico, análogo às estruturas com as quais mantém uma relação direta, que atua na interface com o nível fônico gradiente.

A gramática fonológica e os sistemas de síntese de fala têm uma história comum na tentativa de formalizar a prosódia por meio de regras linguísticas. Nesse contexto, Monaghan e Ladd (1990) descrevem os sistemas de síntese de fala desenvolvidos em um centro de pesquisa de referência internacional que utilizam uma representação fonológica abstrata para gerar as propriedades entoacionais. Essa concepção teórica viabiliza a avaliação precisa de fenômenos, como a distribuição de acentos e a delimitação de domínios prosódicos, o que permite superar as limitações impostas pela análise acústica isolada. Os resultados obtidos com esse modelo contribuem para o aperfeiçoamento das regras de atribuição de acento e evidenciam que a representação fonológica é fundamental para a modelagem da entoação em

[55] Informações disponíveis em: https://cloud.google.com/text-to-speech?hl=pt-br. Acesso em: 24 de julho de 2023.
[56] Informações disponíveis em: https://cloud.google.com/text-to-speech#benefits. Acesso em: 13 de julho de 2024.

sistemas de síntese de fala.

De acordo com Quené e Kager (1992), em consonância com essa perspectiva, os elementos prosódicos da fala sintética derivam de uma estrutura linguística subjacente. Com base na teoria de Nespor e Vogel (1982, 1986), os autores defendem que a acentuação e a estrutura frasal são definidas por uma hierarquia prosódica abstrata. Essa perspectiva revela que a prosódia é um componente gramatical que tem a capacidade de prever os fenômenos entoacionais e acentuais a partir da estrutura fonológica da língua.

O presente estudo aborda a prosódia da fala sintética, com ênfase na entoação declarativa neutra do português brasileiro. Apesar de a prosódia ser um tema de ampla pesquisa em outras línguas, como os clássicos trabalhos referentes à entoação sintética do inglês elaborados por Pierrehumbert (1981, 1993), a literatura dedicada à prosódia do português brasileiro em um contexto computacional ainda é limitada, especialmente em relação à estrutura entoacional de enunciados declarativos neutros produzidos por modelos contemporâneos de síntese de fala, como o Google Cloud Text-to-Speech. Dessa forma, o objetivo da investigação é caracterizar, sob uma perspectiva fonológica, a entoação declarativa neutra do português brasileiro na fala sintética, a fim de subsidiar os futuros avanços na modelagem da prosódia dessa variedade linguística em sistemas de síntese de fala baseados em IA. Ao reconhecer as pesquisas existentes que abordam a prosódia da fala sintética em outras línguas, o trabalho espera ampliar a discussão a respeito do tema e promover um acréscimo de conhecimento científico acerca da gramática fonológica da língua portuguesa.

Após a explanação dos conceitos e das teorias que fundamentam o estudo, é necessário detalhar a metodologia empregada na investigação da entoação declarativa neutra produzida pelo Google Cloud Text-to-Speech.

3 METODOLOGIA

Esta seção, destinada à metodologia, é dividida em duas partes principais: a primeira se relaciona aos materiais e a segunda, aos métodos.

3.1 MATERIAIS

A fim de investigar a entoação da fala sintética, é organizada uma amostra composta de sete enunciados declarativos neutros.[57] Tais sentenças, exemplificadas em (1), são extraídas do *corpus* elaborado por Tenani (2002).[58]

(1)

1. Batata combina com peixe.
2. A casa ficou bonita.
3. Camelôs atacaram policiais.
4. O menino gostou do presente.
5. Panificadores ganharam a disputa.
6. O vendedor chegou atrasado.
7. A pesquisadora terminou os trabalhos.

Ao elaborar os enunciados acima, Tenani (2002) controla o número de sílabas pretônicas em posição inicial de frase entoacional. Segundo a autora, esses enunciados dispõem de uma a quatro sílabas átonas no início absoluto de frase entoacional, sendo que ora a primeira sílaba átona pertence à palavra morfológica, ora a primeira sílaba átona é um artigo e constitui, junto com a palavra morfológica seguinte, uma palavra fonológica. Ademais, afirma a autora, todos os enunciados constituem apenas uma frase entoacional (coextensiva a um enunciado fonológico), formada por três frases fonológicas não ramificadas, cada qual

[57] Na versão inicial da pesquisa, o enunciado "Comerciantes elegeram seus representantes", também criado por Tenani (2002), integra o *corpus* da investigação. Ainda que ele apresente semelhanças entoacionais com os demais enunciados, independentemente do modelo de voz sintética, como o contorno entoacional pré-nuclear ascendente (L*+H), o contorno entoacional nuclear descendente (H+L* L%) e a alta densidade tonal, decide-se por excluí-lo da amostra. Essa escolha se fundamenta no fato de que o referido enunciado contém quatro palavras fonológicas, enquanto os demais contam com três. Assim, com o objetivo de padronizar a configuração dos constituintes prosódicos, mantêm-se apenas os enunciados formados por três palavras fonológicas.

[58] Mesmo que uma parte dos enunciados contenha consoantes desvozeadas e palavras oxítonas, fatores que podem dificultar a identificação dos tons nas sílabas finais e influenciar a visualização do contorno melódico (Fernandes, 2007b), essas estruturas possibilitam examinar a configuração fonológica da entoação declarativa neutra no português brasileiro de modo adequado aos objetivos da pesquisa, conforme proposto no estudo pioneiro de Tenani (2002) para a fala natural. A aplicação do recurso de interpolação fonética preserva a integridade acústica do contorno entoacional na representação gráfica da melodia e assegura que a notação fonológica registre os fenômenos tonais com precisão, sem introduzir distorções que comprometam a descrição gramatical. As análises confirmam que os resultados obtidos para a fala sintética correspondem integralmente aos da fala natural em relação à estrutura fonológica da entoação para a classe das declarações neutras.

contendo uma palavra fonológica (portadora de um acento primário ou lexical), como é possível observar em (2), adaptado de Tenani (2002, p. 35).

(2)

1. [[[Batata]ω]ϕ [combina]ω]ϕ [com peixe.]ω]ϕ]*I*
2. [[[A casa]ω]ϕ [ficou]ω]ϕ [bonita.]ω]ϕ]*I*
3. [[[Camelôs]ω]ϕ [atacaram]ω]ϕ [policiais.]ω]ϕ]*I*
4. [[[O menino]ω]ϕ [gostou]ω]ϕ [do presente.]ω]ϕ]*I*
5. [[[O vendedor]ω]ϕ [chegou]ω]ϕ [atrasado.]ω]ϕ]*I*
6. [[[Panificadores]ω]ϕ [ganharam]ω]ϕ [a disputa.]ω]ϕ]*I*
7. [[[A pesquisadora]ω]ϕ [terminou]ω]ϕ [os trabalhos.]ω]ϕ]*I*

Os enunciados são gerados pelas vozes Standard (pt-BR-Standard-A, pt-BR-Standard-B e pt-BR-Standard-C), Neural2 (pt-BR-Neural2-A, pt-BR-Neural2-B e pt-BR-Neural2-C) e WaveNet (pt-BR-Wavenet-A, pt-BR-Wavenet-B e pt-BR-Wavenet-C).[59] As variedades de voz A e C são femininas, enquanto a B é masculina. Ao todo, são totalizados 63 dados de fala sintética, com uma velocidade de reprodução (taxa de elocução ou velocidade de fala) normal. A justificativa para o trabalho com as vozes Standard, Neural2 e WaveNet é a oportunidade de o estudo incluir os modelos de voz sintética disponíveis, no produto estudado, para o português brasileiro.[60]

De acordo com o Google Cloud Text-to-Speech, as vozes disponibilizadas por essa tecnologia diferem na forma de produção, uma vez que se fundamentam em distintos modelos de síntese de fala.[61] A voz Standard emprega a conversão paramétrica de texto em fala, em que os algoritmos de processamento de sinais, denominados *vocoders*, geram o áudio. Em um nível mais avançado, a voz Neural2 representa uma categoria *premium* que adota a mesma tecnologia das Vozes Personalizadas, o que possibilita uma síntese vocal sofisticada sem a necessidade de treinamento específico.[62] Essa voz está disponível tanto em *endpoints*

[59] A seleção do Google Cloud Text-to-Speech se deve à relevância do serviço no setor de tecnologia digital. Com foco nesse sistema, torna-se possível realizar uma análise mais detalhada da entoação declarativa neutra no português brasileiro. Para os estudos subsequentes, são recomendadas novas comparações com outros sistemas de síntese de fala.

[60] As vozes Standard, Neural2 e WaveNet representam diferentes gerações de síntese de fala da empresa Google.

[61] As informações sobre os três modelos de síntese de fala correspondem ao conteúdo disponível na página oficial do produto.

[62] A desnecessidade de um treinamento específico para a geração da voz Neural2 implica que a qualidade fonológica observada nesse tipo de modelo de síntese de fala decorre não somente do aprendizado estatístico do sistema, mas também das representações simbólicas implícitas nos dados de fala natural utilizados como entrada. Essas representações, caracterizadas como linguísticas e cognitivas, tendem a refletir os padrões prosódicos das línguas naturais e permitem que a voz sintética reproduza as configurações fonológicas compatíveis com a estrutura gramatical. Assim, entende-se que a eficácia da síntese de fala em níveis mais avançados advém da interação entre os mecanismos de aprendizagem probabilística e os princípios formais da gramática fonológica encontrados nas gravações de fala natural.

internacionais quanto regionais. Ademais, o modelo WaveNet constitui uma abordagem inovadora para a síntese de fala, pois aprimora a naturalidade da voz sintetizada e reproduz, com mais precisão, a entoação e a fluidez da fala humana, além de abranger as variações sutis na pronúncia de segmentos, de sílabas e de palavras.[63]

Ao investigar a entoação declarativa neutra no português brasileiro, Tenani (2002) determina que, para cada um dos enunciados, são realizadas duas leituras por três informantes da mesma faixa etária, sexo, grau de escolaridade e dialeto. A amostra da presente pesquisa, por outro lado, é formada por 63 ocorrências, visto que os sete enunciados declarativos neutros são gerados pelas variedades A, B e C dos modelos de voz Standard, Neural2 e WaveNet, pertencentes ao mesmo sistema de síntese de fala. Conforme é possível observar na próxima seção, essa quantidade de dados é suficiente para descrever as características entoacionais do conjunto de enunciados selecionados, com base nos objetivos estabelecidos pelo pesquisador para o desenvolvimento do trabalho e na evidente semelhança detectada entre as ocorrências investigadas, em termos de estrutura entoacional. Desse modo, não se justifica a ampliação do número de dados, haja vista que eles se demonstram adequados para o cumprimento dos objetivos da pesquisa.[64]

A pesquisa não desconsidera a complexidade do Google Cloud Text-to-Speech, tampouco tem a intenção de realizar uma análise técnica detalhada da arquitetura e do funcionamento interno desse sistema. Em vez disso, a abordagem interdisciplinar do trabalho se concentra no diálogo entre a Linguística e a Ciência da Computação, por meio da análise dos padrões entoacionais da fala sintética em enunciados declarativos neutros. A investigação reconhece as principais abordagens utilizadas na construção de sistemas de síntese de fala, como as diferenças entre os modelos concatenativos, paramétricos e articulatórios, apresentados na fundamentação teórica, além da evolução da tecnologia de fala desde os primeiros modelos de síntese até as redes neurais artificiais. A seleção das modalidades de voz oferecidas pelo Google Cloud Text-to-Speech evidencia uma preocupação em abranger os distintos níveis de sofisticação tecnológica disponibilizados pelo sistema de síntese de fala estudado, com a premissa confirmada de que eles geram uma estrutura melódica análoga à fala natural para a declaração neutra. Sendo assim, o estudo não somente descreve a entoação

[63] Informações disponíveis em: https://cloud.google.com/text-to-speech/docs/wavenet?hl=pt-br. Acesso em: 8 de setembro de 2023.

[64] Com o auxílio de Carlos Elísio Nascimento da Silva, a quem o pesquisador agradece a valiosa contribuição, são realizados alguns testes estatísticos para a avaliação do tamanho da amostra da pesquisa. Para tanto, utiliza-se a biblioteca NumPy, da linguagem de programação Python, por meio da ferramenta *Data Analysis with ChatGPT* (https://chatgpt.com/g/g-HMNcP6w7d-data-analyst; acesso em: 21 de abril de 2024). Os resultados indicam a suficiência da amostra utilizada para os achados observados, pois apresenta uma forte tendência estatística e tamanhos de efeito elevados, o que dispensa a necessidade de ampliação do material.

declarativa neutra da fala produzida pelo Google Cloud Text-to-Speech, mas também reflete sobre o progresso linguístico-computacional da tecnologia, particularmente no que se refere à modelagem da entoação de um conjunto de enunciados do português brasileiro.

A descrição das diferentes vozes do Google Cloud Text-to-Speech contextualiza, para o leitor, a tecnologia de síntese de fala utilizada na pesquisa, sem o estabelecimento de uma diferenciação de desempenho técnico de uma voz em relação às outras. Embora as vozes Standard, Neural2 e WaveNet empreguem distintas estratégias de geração de fala, o foco do estudo não é uma avaliação comparativa entre elas, mas a caracterização da entoação dos enunciados declarativos neutros sintéticos como um todo, em relação aos padrões estabelecidos para a fala natural do português brasileiro. Ainda que o trabalho não se concentre na variação técnica de cada voz, ele contribui para a discussão sobre a modelagem prosódica na fala sintética.

No momento da concepção da pesquisa, as variedades de voz disponíveis são A, B e C para os três modelos de voz (Standard, Neural2 e WaveNet). Entretanto, após a atualização da ferramenta, toma-se a decisão de não incluir as novas variedades, visto que o escopo do estudo se encontra nas questões fonológicas. Embora as novas variedades apresentem características fonéticas adicionais, como diferentes qualidades de voz, hipotetiza-se que, no nível fonológico, a estrutura entoacional dos enunciados declarativos neutros permanece inalterada. O estudo demonstra uma padronização fonológica do contorno entoacional dos enunciados, com base nas variedades A, B e C. Os tópicos exclusivamente relacionados a aspectos idiossincráticos ou sociolinguísticos não são objeto da pesquisa, que se concentra na análise fonológica do contorno de entoação, de acordo com a perspectiva teórica escolhida.

Não é elaborada qualquer hipótese com relação à possível diferença técnica e acústica no modelo (Standard, Neural2 e WaveNet) e na variedade (A, B e C) de voz. Como o resultado da pesquisa é uma interpretação fonológica do fenômeno, o propósito da investigação é determinar as características linguísticas gerais da entoação declarativa neutra na fala sintética no português brasileiro. Independentemente da expressividade das vozes analisadas, o objetivo da pesquisa é identificar os padrões entoacionais dos enunciados examinados. Em outras palavras, a intenção do trabalho é obter informações sobre a organização fonológica da estrutura melódica. A saída acústica gerada pelo Google Cloud Text-to-Speech pode variar, por exemplo, em função de fatores idiossincráticos das vozes e dos mecanismos computacionais envolvidos na composição do sistema de síntese de fala. Esses aspectos não são abordados na pesquisa, uma vez que ela se concentra, como explicado acima, na caracterização fonológica do contorno melódico declarativo neutro produzido pela

ferramenta estudada, com o intuito de fornecer implicações linguísticas à prosódia do português brasileiro.

No que diz respeito às características estruturais das ocorrências, é imperativo observar que, embora os enunciados declarativos neutros elaborados por Tenani (2002) configurem períodos sintáticos simples, ou seja, formados por apenas um verbo, eles dispõem de propriedades gramaticais que os diferenciam entre si.[65] Dessa forma, a amostra selecionada também é suficiente para investigar as características entoacionais de enunciados declarativos neutros sintéticos no contexto de períodos sintáticos simples, já que as estruturas linguísticas das sentenças do *corpus* diferem em termos morfológicos e sintáticos, conforme detalhado no Quadro 2.[66] Mesmo com os constituintes organizados segundo o padrão básico do português (Pezatti, 1992, 2014; Castilho, 2019), a variação estrutural nas categorias morfológicas e nas funções sintáticas mostra que cada enunciado apresenta um significado específico, conforme os fatores gramaticais envolvidos na composição das sentenças.[67] Outro argumento favorável à escolha dessa amostra é o fato de que a pesquisa de Tenani (2002), uma bibliografia essencial para os estudos prosódicos e constantemente citada desde a defesa do trabalho, dispõe de resultados robustos na investigação da estrutura entoacional da declaração neutra no português brasileiro, ao adotar os mesmos enunciados para a análise da fala natural.

[65] Ainda que outras estruturas sejam passíveis de estudo, a decisão de restringir o *corpus* a enunciados declarativos neutros de períodos sintáticos simples também se baseia na necessidade de controle das variáveis prosódicas e de garantia do rigor teórico e metodológico dentro do período regulamentar da pesquisa.

[66] A análise morfológica e sintática é realizada exclusivamente pelo pesquisador. O uso do ChatGPT/OpenAI (GPT-4o mini) serve apenas para organizar o quadro em linhas e colunas.

[67] O exame das relações gramaticais tende a considerar as funções de sujeito e de predicado ou de sujeito, de verbo e de complemento, além de abranger a análise da ordenação desses constituintes na sentença (Kato; Mioto, 2020).

Quadro 2 – Aspectos morfológicos e sintáticos dos enunciados declarativos neutros

Enunciados	**Aspectos morfológicos**	**Aspectos sintáticos** **(continua)**
Batata combina com peixe.	"Batata": substantivo feminino singular; "combina": verbo transitivo indireto, 3ª pessoa do singular, presente do indicativo; "com": preposição; "peixe": substantivo masculino singular.	"Batata" (sujeito simples - sintagma nominal) + "combina com peixe" (predicado verbal): - "combina" (núcleo do predicado - sintagma verbal); - "com peixe" (sintagma preposicional - objeto indireto).
A casa ficou bonita.	"A": artigo definido feminino singular; "casa": substantivo feminino singular; "ficou": verbo de ligação, 3ª pessoa do singular, pretérito perfeito do indicativo; "bonita": adjetivo qualificativo feminino singular	"A casa" (sujeito simples - sintagma nominal) + "ficou bonita" (predicado nominal): - "ficou" (núcleo do predicado - sintagma verbal); - "bonita" (predicativo do sujeito - sintagma adjetival).
Camelôs atacaram policiais.	"Camelôs": substantivo masculino plural; "atacaram": verbo transitivo direto, 3ª pessoa do plural, pretérito perfeito do indicativo; "policiais": substantivo masculino plural.	"Camelôs" (sujeito simples - sintagma nominal) + "atacaram policiais" (predicado verbal): - "atacaram" (núcleo do predicado - sintagma verbal); - "policiais" (objeto direto - sintagma nominal).
O menino gostou do presente.	"O": artigo definido masculino singular; "menino": substantivo masculino singular; "gostou": verbo transitivo indireto, 3ª pessoa do singular, pretérito perfeito do indicativo; "do": contração da preposição "de" com o artigo "o"; "presente": substantivo masculino singular.	"O menino" (sujeito simples - sintagma nominal) + "gostou do presente" (predicado verbal): - "gostou" (núcleo do predicado - sintagma verbal); - "do presente" (sintagma preposicional - objeto indireto).
O vendedor chegou atrasado.	"O": artigo definido masculino singular; "vendedor": substantivo masculino singular; "chegou": verbo intransitivo, 3ª pessoa do singular, pretérito perfeito do indicativo; "atrasado": adjetivo qualificativo masculino singular.	"O vendedor" (sujeito simples - sintagma nominal) + "chegou atrasado" (predicado verbo-nominal): - "chegou" (núcleo verbal do predicado - sintagma verbal); - "atrasado" (núcleo nominal do predicado - predicativo do sujeito - sintagma adjetival).

Quadro 2 – Aspectos morfológicos e sintáticos dos enunciados declarativos neutros

Enunciados	**Aspectos morfológicos**	**Aspectos sintáticos** **(conclusão)**
Panificadores ganharam a disputa.	“Panificadores”: substantivo masculino plural; “ganharam”: verbo transitivo direto, 3ª pessoa do plural, pretérito perfeito do indicativo; “a”: artigo definido feminino singular; “disputa”: substantivo feminino singular.	“Panificadores” (sujeito - sintagma nominal) + “ganharam a disputa” (predicado verbal): - “ganharam” (núcleo do predicado - sintagma verbal); - “a disputa” (objeto direto - sintagma nominal).
A pesquisadora terminou os trabalhos.	“A”: artigo definido feminino singular; “pesquisadora”: substantivo feminino singular; “terminou”: verbo transitivo direto, 3ª pessoa do singular, pretérito perfeito do indicativo; “os”: artigo definido masculino plural; “trabalhos”: substantivo masculino plural.	“A pesquisadora” (sujeito simples - sintagma nominal) + “terminou os trabalhos” (predicado verbal): - “terminou” (núcleo do predicado - sintagma verbal); - “os trabalhos” (objeto direto - sintagma nominal).

Fonte: Elaborado pelo autor (2025) com o auxílio do ChatGPT/OpenAI (GPT-4o mini).[68]

Apesar de a escrita não conseguir representar a totalidade da variação melódica da fala (Oliveira Jr., 2022), os sinais de pontuação desempenham um papel pertinente na sinalização da modalidade do enunciado, ao indicarem, dentre outros fenômenos linguístico-discursivos, a distinção, no português, entre as frases declarativas, marcadas pelo ponto final, e as frases interrogativas, indicadas pelo ponto de interrogação (Cagliari, 1989; Pacheco, 2003; Galdino; Silva, K.; Oliveira Jr., 2021; Moraes; Rilliard, 2022). Um aspecto metodológico que requer atenção se refere à configuração do Google Cloud Text-to-Speech, o qual não permite especificar a modalidade dos enunciados por meio de opções computacionais explícitas. Ao concluir os enunciados com um ponto final e considerar a estrutura morfológica e sintática das sentenças, o sistema infere que a realização fonética deve se ajustar, em uma relação entre a fala e a escrita, à entoação descendente típica de frases declarativas (Cagliari, 1989, 2007; Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Pacheco, 2003; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). Quanto à neutralidade, os enunciados declarativos não são inseridos no sistema com o uso de letras maiúsculas (exceto pela inicial da sentença),

[68] Disponível em: https://openai.com/. Acesso em: 10 de setembro de 2024.

nem com outras mudanças gráfico-visuais, a fim de evitar, na concepção do pesquisador, que a máquina atribua, no processo de síntese de fala, alguma proeminência adicional a elementos fonológicos que não correspondam ao contorno entoacional nuclear.

O pesquisador responsável pelo desenvolvimento da pesquisa opta pelo trabalho com enunciados declarativos neutros constituídos de períodos sintáticos simples, haja vista que uma investigação com períodos sintáticos compostos demandaria mais tempo. A depender do tipo e da subcategoria das estruturas complexas, há diferentes aspectos gramaticais envolvidos na formação das sentenças, o que leva a distinções nas características prosódicas dos enunciados. Em pesquisas anteriores do autor (Tojeira-Ramos; Pezatti, 2021, 2022; Tojeira-Ramos, 2024; Tojeira-Ramos; Guiraldelli, 2024), verifica-se que, em períodos sintáticos compostos, as estruturas gramaticais apresentam diferentes propriedades prosódicas, a depender de fatores pragmáticos, semânticos e morfossintáticos. Sendo assim, o trabalho com períodos sintáticos compostos, em dados de fala sintética, não é realizado durante a execução da investigação e, portanto, pode ser incluído na agenda de pesquisa como uma sugestão para estudos futuros.

Assim como no trabalho de Kato e Nascimento (2020), o objeto de estudo é delimitado ao desempenho linguístico observável em um *corpus*, compreendido como uma manifestação da chamada Língua-E (Chomsky, 1986). O *corpus* da pesquisa consiste em enunciados declarativos neutros produzidos computacionalmente a partir de dados de fala humana. Apesar de a tecnologia de síntese de fala não dispor de uma faculdade da linguagem própria, a geração estatística de enunciados projeta, de forma mediada, a complexidade do sistema linguístico internalizado pelos falantes que fornecem os dados naturais de treinamento. Dessa maneira, ao analisar os contornos melódicos das vozes sintéticas para as declarações neutras, é possível investigar, ainda que de modo indireto, como os eventos tonais e os padrões entoacionais do português brasileiro se relacionam ao funcionamento da gramática humana.

Embora o foco do estudo seja a língua registrada em um *corpus*, conforme indicam Kato e Nascimento (2020), torna-se necessário reconhecer que os dados de fala humana que instruem os sistemas computacionais inevitavelmente remetem à Língua-I, ou seja, à gramática internalizada (Chomsky, 1999). O desempenho observado na fala sintética não representa uma faculdade específica da linguagem da máquina, mas uma projeção estatística das regularidades linguísticas dos seres humanos. Com base nos autores, a metodologia adotada na pesquisa articula a análise acústico-fonológica dos enunciados sintéticos com a intuição do pesquisador, acessada por meio da inspeção auditiva, e as descrições teóricas e empíricas da literatura (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Cagliari, 2007;

Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), a fim de verificar em que medida os contornos entoacionais produzidos computacionalmente correspondem às configurações melódicas do português brasileiro para os enunciados declarativos neutros.

A próxima subseção é reservada à exposição dos métodos da pesquisa.

3.2 MÉTODOS

A pesquisa considera que a descrição e a interpretação da entoação dependem tanto da análise detalhada de parâmetros fonéticos quanto da identificação de categorias e de estruturas fonológicas que organizam os fenômenos melódicos no sistema linguístico. Nesse sentido, é estabelecido um diálogo entre a Fonética e a Fonologia, em que os critérios fonéticos constituem a base instrumental para a transcrição fonológica da entoação declarativa neutra.

Após a coleta dos dados da pesquisa, os métodos utilizados para descrever a estrutura entoacional dos arquivos de fala sintética e compará-los com a fala natural são organizados em três etapas complementares. Para fins didáticos, cada uma é apresentada e detalhada em subseções específicas.[69]

3.2.1 Identificação dos eventos tonais nos enunciados declarativos neutros

A primeira etapa diz respeito à identificação dos eventos tonais associados ao contorno de entoação dos enunciados declarativos neutros. Antes do tratamento acústico dos dados, é realizada uma inspeção auditiva das ocorrências coletadas.[70] Essa conduta se alinha ao posicionamento de Cagliari (2012a), segundo o qual, caso o resultado desejado seja uma interpretação linguística do fenômeno, a análise física não pode prescindir de uma escuta das ocorrências, visto que, conforme apontado adiante pelo pesquisador, "o ouvido mais o cérebro

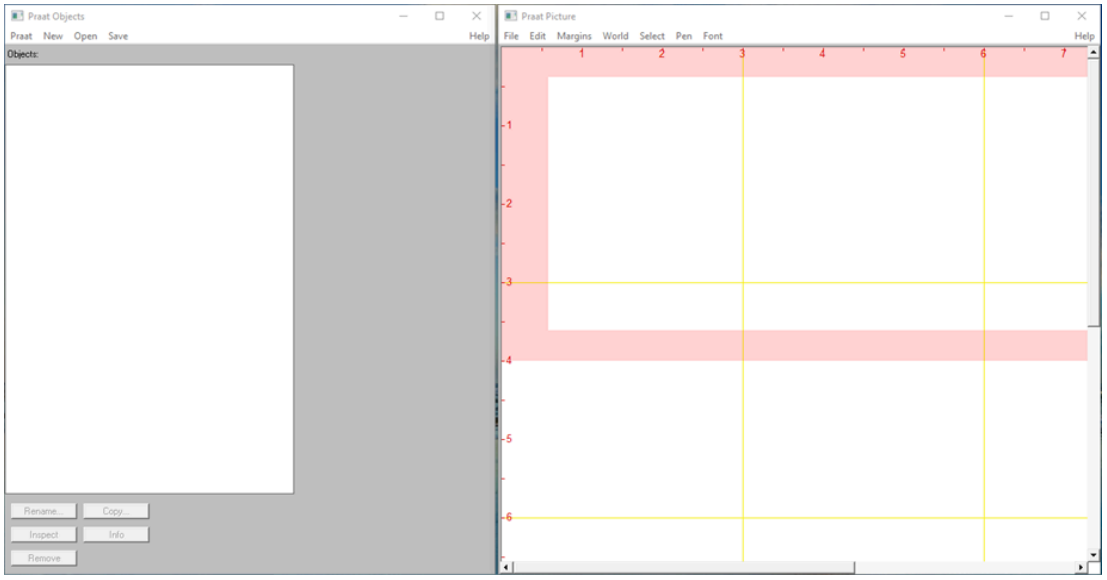
[69] A definição dos parâmetros analisados decorre da necessidade de uma articulação entre a dimensão fonética e a dimensão fonológica, para que seja possível interpretar os traços entoacionais como elementos constitutivos do sistema linguístico. Essa integração propicia a coerência entre a descrição empírica e a representação teórica da entoação, além de elucidar os aspectos físicos da fala e as regularidades simbólicas que organizam o contorno melódico dos enunciados declarativos neutros.

[70] O leitor deve se lembrar de que a "inspeção auditiva" se refere à escuta analítica do pesquisador, distinta da "análise auditiva", em que os julgamentos são realizados por participantes humanos. Na pesquisa, a inspeção auditiva atua como um método auxiliar para detectar os padrões perceptualmente salientes, que orientam a confirmação do linguista por meio da análise acústica.

e a mente do indivíduo formam um laboratório acústico altamente sofisticado para a percepção de sons da fala" (Cagliari, 2012a, p. 10).[71]

Além da inspeção auditiva (audição dos arquivos de fala sintética),[72] é adotado o programa computacional (*software*) Praat (versão 6.4.04).[73] O referido programa é criado por Paul Boersma e David Weenink, pertencentes ao Instituto de Ciências Fonéticas, da Universidade de Amsterdã (Holanda do Norte, Países Baixos).[74] A qualidade da resolução no tratamento do sinal acústico e a gratuidade (Silva, A., 2020a) justificam a escolha do Praat para a realização da pesquisa. Na Figura 25, são exibidas as janelas Praat Objects (à esquerda) e Praat Picture (à direita). A primeira janela (Praat Objects) é usada para abrir (em formato .wav), anotar (com o auxílio do TextGrid) e analisar (de um ponto de vista acústico) os arquivos de fala sintética, enquanto a segunda janela (Praat Picture) é empregada para gerar as imagens dos contornos entoacionais.[75]

Figura 25 – Janelas Praat Objects e Praat Picture



Fonte: Elaborado pelo autor (2025).

[71] A ausência de um experimento de percepção entoacional na pesquisa é justificada pela limitação de tempo e de escopo. O estudo tem como prioridade estabelecer uma comparação entre a fala natural e a sintética, garantir uma descrição robusta da estrutura entoacional e realizar um mapeamento entre os níveis fonético e fonológico. A aplicação de testes de percepção requer um desenho experimental distinto, o que excede o escopo e o cronograma da investigação. Sugere-se, portanto, um estudo psicoacústico complementar, que contraste vozes naturais e sintéticas, para verificar a identificação dos contornos melódicos e a naturalidade percebida. Tal investigação compõe o escopo do Doutorado do autor, a ser desenvolvido a partir de 2026.

[72] É preciso sublinhar que o diálogo com a síntese de fala contemporânea não deve se restringir à inspeção auditiva da naturalidade. Também é necessário examinar como as estruturas prosódicas são processadas e reproduzidas. Dessa forma, o estudo utiliza os instrumentos linguísticos clássicos para uma leitura crítica das saídas acústicas da IA.

[73] O *software* Praat se encontra disponível, para *download*, em: http://www.fon.hum.uva.nl/praat/. Acesso em: 20 de janeiro de 2024.

[74] Instruções básicas de como utilizar o *software* Praat são encontradas em Cristófaro Silva (2021a).

[75] Instruções básicas de como anotar e segmentar arquivos sonoros são encontradas em Cristófaro Silva (2021b).

De acordo com Constantini e Barbosa (2015), diversos fatores podem interferir na extração de medidas segmentais e suprassegmentais em gravações de fala durante a identificação do locutor, como o disfarce de voz, a distorção de sinal e o ruído de fundo. Contudo, tais questões não preocupam a pesquisa, pois os dados de fala são obtidos sem interferências, diretamente do computador, a partir de uma tecnologia de síntese de fala.

Silva, A. (2010, 2020a) destaca que a competência em utilizar o Praat não é sinônimo de habilidade em realizar uma pesquisa em Fonética Acústica. Dessa forma, segundo a autora, é imprescindível ter clareza sobre o que se pretende analisar no sinal acústico e quais parâmetros físicos são mais apropriados. Quando o objetivo é verificar o contorno entoacional de um enunciado, como no caso da pesquisa, é necessário observar a curva da F_0.

Conforme mencionado por Silva, A. (2010, 2020a), além de definir os parâmetros de análise a partir dos aspectos que se deseja verificar na cadeia de fala, cabe ao pesquisador interpretar os dados obtidos. A interpretação dos resultados é fundamental para confirmar ou refutar uma hipótese formulada. A autora enfatiza que essa tarefa exige muito mais do que saber usar o programa. Dessa maneira, embora o conhecimento do uso do Praat seja um requisito instrumental para a realização de uma pesquisa em Fonética, ele não é suficiente para a elaboração de um estudo fonético propriamente dito.[76]

Interessa ao trabalho a variação fonética da F_0, que é observada com a utilização dos critérios adotados por Galvão Passetti (2021).[77] Esses critérios, também utilizados por Tojeira-Ramos e Pezatti (2022), Tojeira-Ramos (2024) e Tojeira-Ramos e Guiraldelli (2024), são descritos nos três próximos parágrafos. Destaca-se, portanto, a relevância do exame dos detalhes fonéticos para a identificação mais precisa das características fonológicas dos contornos entoacionais (Moraes, 2016).[78]

---

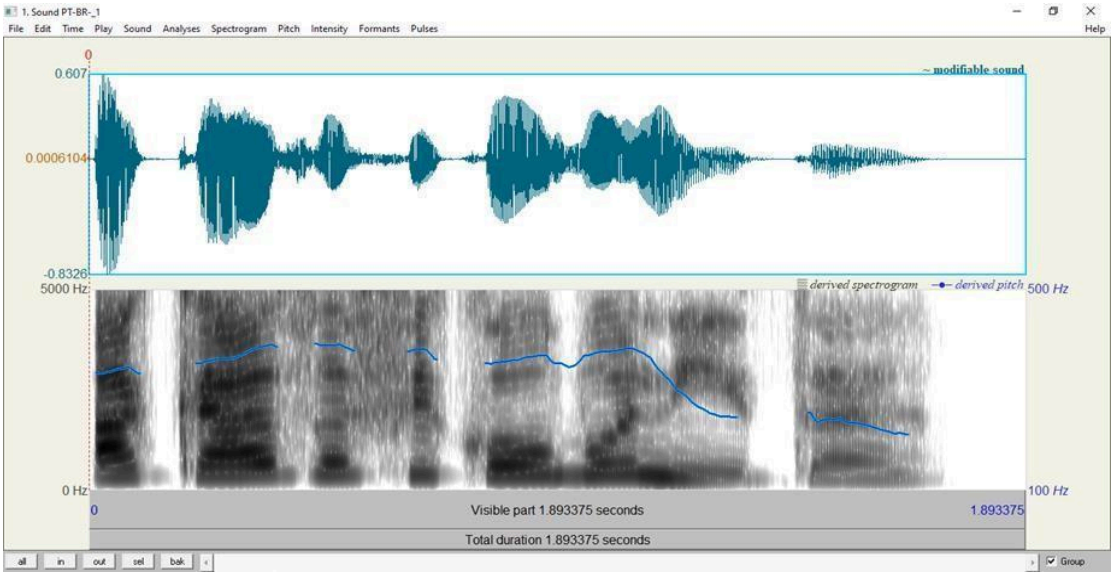
[76] A análise é feita com o uso de técnicas da Fonética Acústica para a identificação das propriedades prosódicas que estruturam a entoação declarativa neutra do português brasileiro no nível fonológico. A interpretação fonológica orienta essa análise e explica como a prosódia se integra ao sistema linguístico.

[77] Em concordância com Ladd (2008), uma descrição fonológica completa não se limita a fórmulas abstratas, mas também especifica como essas fórmulas se realizam, ou seja, descreve o mapeamento dos elementos fonológicos categóricos para os parâmetros acústicos contínuos. Nesse sentido, de acordo com o autor, a Fonologia Entoacional Autossegmental e Métrica tem o objetivo fonológico de caracterizar os contornos entoacionais por meio de uma sequência de elementos categoricamente distintos e o objetivo fonético de fornecer um mapeamento dos elementos fonológicos para os parâmetros acústicos contínuos. Trata-se, portanto, de mais um argumento que ratifica o diálogo entre a Fonética e a Fonologia.

[78] Os detalhes fonéticos da variação melódica são considerados de acordo com a relevância gramatical. O tratamento do sinal acústico torna viável a distinção entre as propriedades gramaticalmente significativas e as variações irrelevantes para o sistema linguístico. Além disso, os detalhes fonéticos constituem a base para a recuperação de categorias entoacionais. O processo de análise identifica, em primeiro lugar, o conjunto mínimo de indícios utilizados pelo ouvido humano para diferenciar as categorias simbólicas e, em seguida, remove as microvariações fonéticas que não alteram a classificação melódica. O uso de técnicas de tratamento acústico tem como finalidade identificar as propriedades que são funcionalmente perceptíveis e reduzir as alterações sem qualquer pertinência gramatical para a declaração neutra.

Quanto ao primeiro critério, para a filtragem do contorno fonético da F_0, são definidos os intervalos (*pitch range*) de busca de valores (*pitch settings*) entre 100 e 500 Hz para as vozes femininas (mais agudas) e entre 75 e 300 Hz para as vozes masculinas (mais graves).[79] Esses parâmetros seguem as orientações do manual de uso do *software* Praat, escrito por Paul Boersma e David Weenink, vigentes na época da consulta realizada por Galvão Passetti (2021).[80] Nas Figuras 26 e 27, encontram-se o oscilograma (parte superior) e o espectrograma (parte inferior) de arquivos sonoros cujos intervalos de busca de valores são definidos entre 100 e 500 Hz (voz feminina) e entre 75 e 300 Hz (voz masculina). Segundo Cristófaro Silva *et al.* (2019), o oscilograma (ou forma de onda) registra, no eixo das abscissas (horizontal), os instantes de tempo, medidos em segundos, e, no eixo das ordenadas (vertical), os parâmetros de amplitude que constituem a onda sonora, medidos em decibéis (dB). Já o espectrograma (ou sonograma) registra, no eixo das abscissas, os instantes de tempo, medidos em segundos (s), e, no eixo das ordenadas, as frequências que constituem a onda sonora, medidas em Hz (Cristófaro Silva *et al.*, 2019). Observa-se, nos espectrogramas das Figuras 26 e 27, um traçado azul, que sinaliza a variação fonética da F_0.

Figura 26 – Oscilograma e espectrograma do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Standard-A

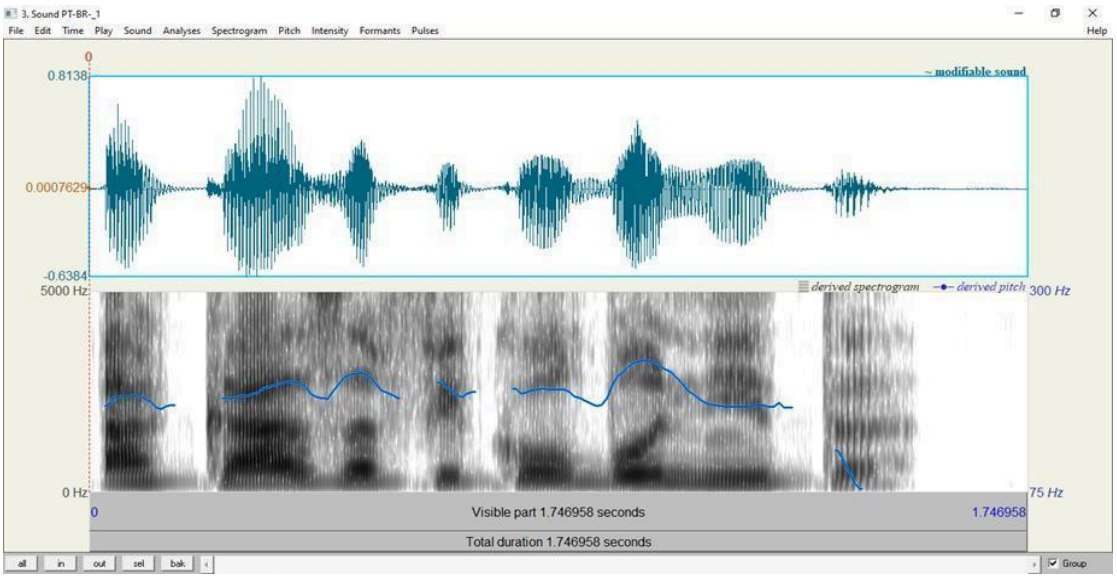


Fonte: Elaborado pelo autor (2025).

[79] A definição dos intervalos de busca de valores para a F_0 é estabelecida com base em propriedades fisiológicas e acústicas da fala, que reconhecem a influência da massa, da tensão e do comprimento das pregas vocais sobre a taxa de vibração glotal. Essas características estruturais propiciam a ocorrência de variações regulares nos valores da F_0 entre os grupos de falantes: as vozes femininas tendem a apresentar frequências mais altas, em função das pregas vocais mais curtas e leves, enquanto as masculinas, que são mais espessas e pesadas, produzem frequências mais baixas (Cristófaro Silva *et al.*, 2019).

[80] A consulta ao manual de uso do *software* Praat é realizada por Galvão Passetti em abril de 2019.

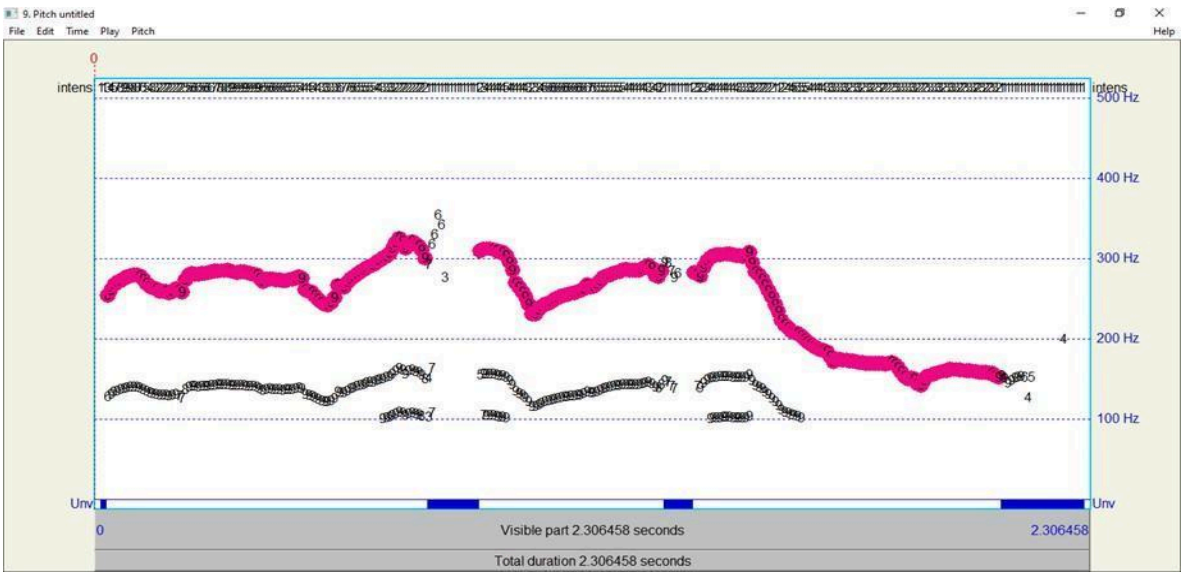
Figura 27 – Oscilograma e espectrograma do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Neural2-B



Fonte: Elaborado pelo autor (2025).

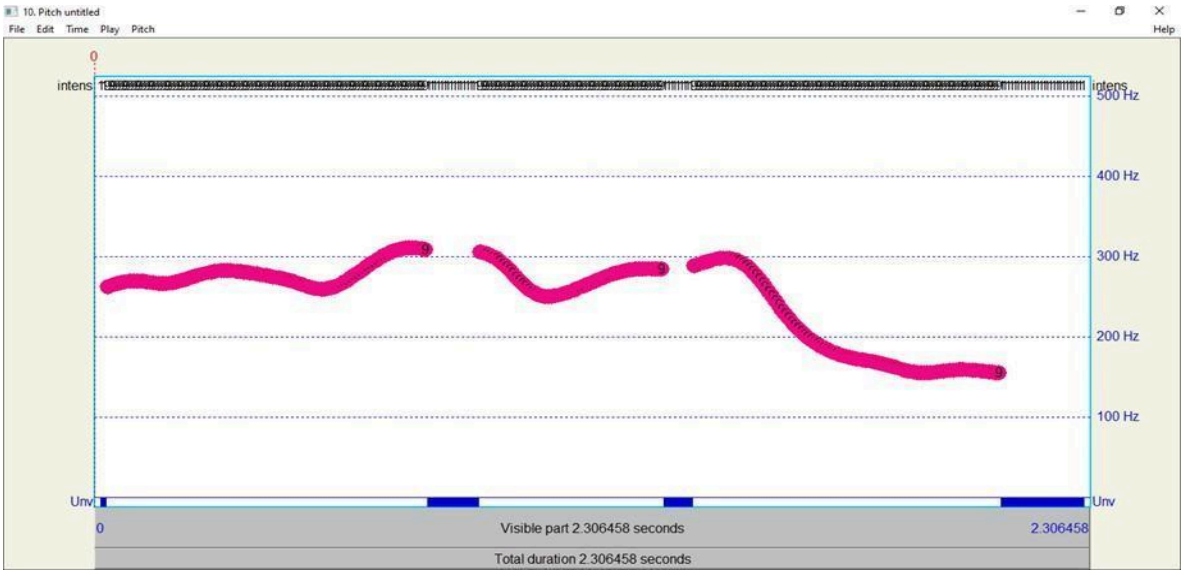
Já em relação ao segundo critério, a F_0 é suavizada (*smooth*) por meio de um filtro (*bandwidth*) de 5 Hz, cujo propósito é realçar as variações melódicas vinculadas à estrutura linguística do enunciado, em consonância com o procedimento empregado por Barbosa e Silva, W. (2012) e referido por Galvão Passetti (2021). Nas Figuras 28 e 29, é exibida a variação da F_0 ao longo do enunciado “O vendedor chegou atrasado”. A primeira figura apresenta, no decorrer do contorno entoacional, alterações micromelódicas, que, além de não servirem a uma função linguística, costumam não ser percebidas pelas estruturas anatomofisiológicas do ouvido humano (Barbosa, 2019, 2022). Na segunda figura, por sua vez, a F_0 é suavizada para ressaltar os movimentos tonais que expressam a estrutura entoacional característica da língua. À vista disso, opta-se por suavizar a F_0 de todos os arquivos sonoros, conforme a exemplificação na Figura 29, a fim de que sejam desconsideradas as variações melódicas desprovidas de implicações para o funcionamento do sistema fonológico da língua portuguesa, o qual é composto da “distinção de sons que os falantes percebem e produzem” (Berti, 2017, p. 2).

Figura 28 – F_0 do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Standard-A, sem suavização



Fonte: Elaborado pelo autor (2025).

Figura 29 – F_0 do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Standard-A, com suavização



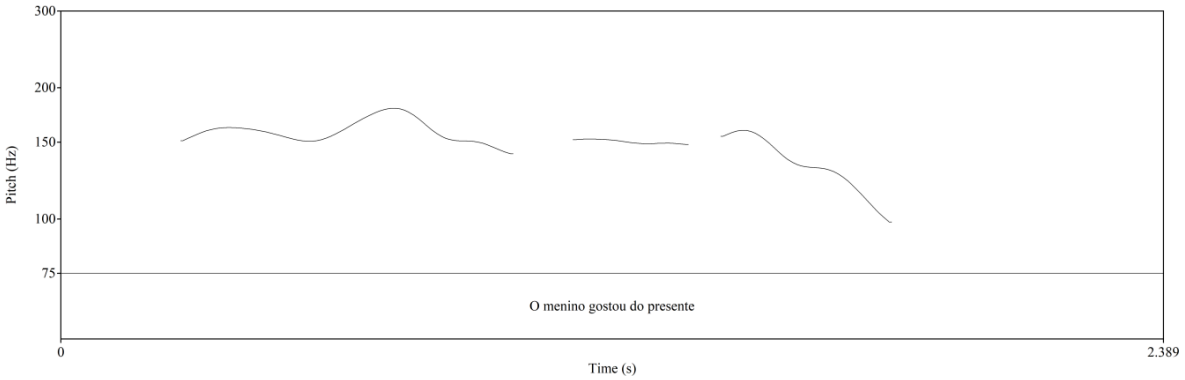
Fonte: Elaborado pelo autor (2025).

O terceiro critério, por outro lado, diz respeito à plotagem dos gráficos da F_0. Para tanto, utiliza-se uma escala logarítmica em Hz no eixo das ordenadas. Essa escolha se justifica pelo fato de que a escala logarítmica aproxima o arranjo gráfico da F_0 à percepção do ouvido humano,[81] segundo o trabalho de Nolan (2003), conforme tomado por Galvão Passetti (2021).

[81] A inspeção auditiva é necessária por causa da importância do padrão sonoro percebido pelo ouvido na comunicação oral (Cagliari, 2012a). Quando associada à análise acústica, ela assegura a confiabilidade dos resultados. Para ratificar as conclusões, propõe-se a realização de futuros experimentos psicoacústicos, com o

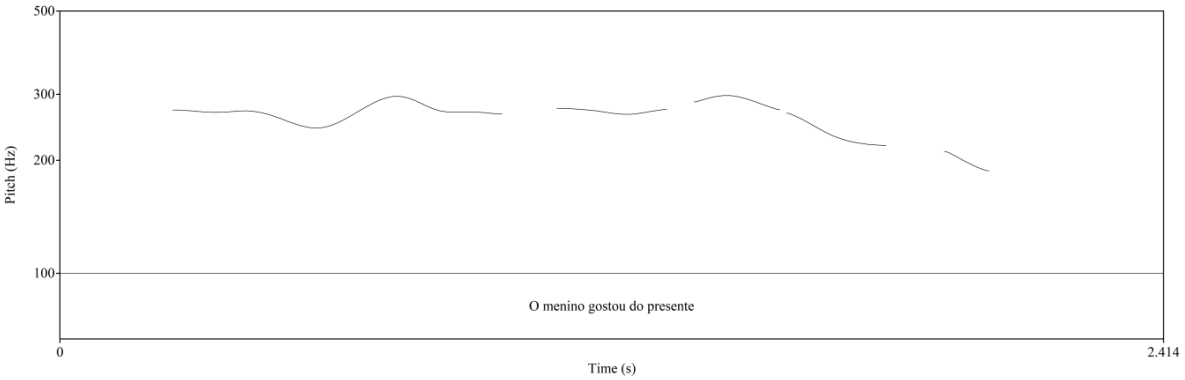
Nas Figuras 30 e 31, são ilustrados dois gráficos concernentes à F_0 do enunciado “O menino gostou do presente”, produzido por vozes sintéticas distintas, com base nos três critérios já descritos.

Figura 30 – F_0 do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Neural2-B, sem interpolação



Fonte: Elaborado pelo autor (2025).

Figura 31 – F_0 do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Neural2-C, sem interpolação



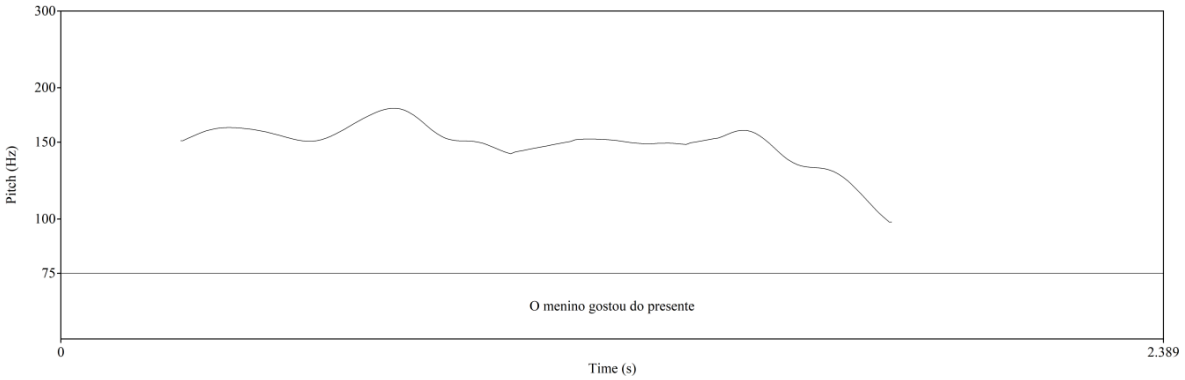
Fonte: Elaborado pelo autor (2025).

Os critérios descritos por Galvão Passetti (2021) são complementados pelo uso do recurso de interpolação fonética (*interpolate*), também empregado por Barbosa e Silva, W. (2012), com a finalidade de tornar o contorno da F_0 mais compreensível, sem que haja lacunas na variação melódica. Quando se considera o fato de que, na Fonologia Entoacional Autossegmental e Métrica, o contorno melódico, formado por uma combinação de acentos tonais e de tons de fronteira, é obtido por meio de interpolações fonéticas entre cada um dos pontos tonais (Collischonn, 2010), torna-se oportuna a utilização dessa técnica na análise acústica dos dados, para evitar possíveis interrupções na visualização do contorno de

objetivo de verificar se os ouvintes humanos reconhecem as categorias gramaticais observadas.

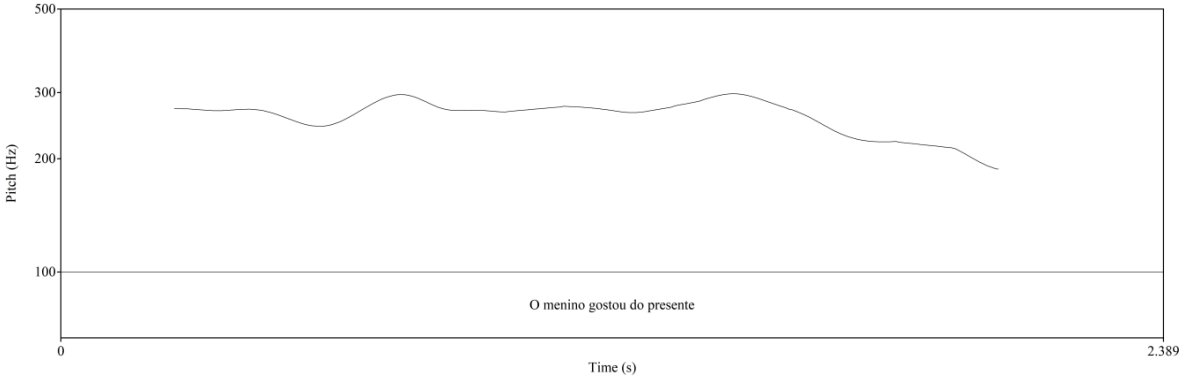
entoação. As lacunas no contorno entoacional sem a interpolação fonética tendem a corresponder aos locais em que são proferidos os segmentos desvozeados (Barbosa; Silva, W., 2012), como determinadas consoantes. No instante em que esses sons são produzidos, não se gera a F_0, em razão da ausência de vibração das pregas vocais (Callou; Leite, 1994; Cagliari, 2007; Massini-Cagliari; Cagliari, 2012; Cristófaro Silva, 2002), o que pode prejudicar a inspeção visual da estrutura melódica do enunciado (Fernandes, 2007b). Nas Figuras 32 e 33, é aplicado o recurso da interpolação (quarto critério) à F_0 do enunciado "O menino gostou do presente". Verifica-se que as curvas da F_0 representadas nos gráficos das Figuras 32 e 33 são, em comparação com as ilustradas nas Figuras 30 e 31, mais evidentes no tocante ao reconhecimento do padrão entoacional dos enunciados.

Figura 32 – F_0 do enunciado "O menino gostou do presente", gerado pela voz pt-BR-Neural2-B, com interpolação



Fonte: Elaborado pelo autor (2025).

Figura 33 – F_0 do enunciado "O menino gostou do presente", gerado pela voz pt-BR-Neural2-C, com interpolação



Fonte: Elaborado pelo autor (2025).

Os eventos tonais associados ao contorno de entoação dos enunciados declarativos neutros são identificados de acordo com o sistema P-ToBI, desenvolvido por Frota, Oliveira,

P., Cruz e Vigário (2015) e pautado nos pressupostos teóricos da Fonologia Entoacional Autossegmental e Métrica e da Fonologia Prosódica. Inicialmente, o ToBI (*Tones and Break Indices*), do qual a adaptação para o português se origina, é criado com base em trabalhos como os de Silverman *et al.* (1992) e Pitrelli, Beckman e Hirschberg (1994), com implicações para a síntese de fala, o que corrobora o uso do sistema em pesquisas relacionadas à entoação sintética. Prioriza-se, no trabalho, a transcrição dos eventos tonais atribuídos à primeira e à última sílabas tônicas e à fronteira direita da frase entoacional, haja vista que eles carreiam informações linguísticas importantes para a identificação da modalidade do enunciado.[82] No entanto, em determinadas análises tonais, são transcritos, para exemplificação, todos os tons associados às palavras fonológicas "cabeça" das frases fonológicas que constituem o contorno de entoação dos enunciados declarativos neutros, conforme realizado no trabalho de Tenani (2002).

A opção pelo P-ToBI se deve à ampla aceitação desse sistema na literatura prosódica sobre a língua portuguesa, como nos trabalhos de Frota *et al.* (2015), Frota e Moraes (2016) e Fernandes-Svartman (2024a, 2024b). Ele permite a representação detalhada da estrutura entoacional por meio da análise dos eventos tonais, como os acentos tonais e os tons de fronteira. Tal sistema é relevante na síntese de fala, pois oferece uma representação que possibilita avaliar, no tocante à declaração neutra, a fidelidade entoacional da fala gerada por algoritmos de IA e compará-la com a fala natural.

A definição da unidade portadora de proeminência melódica na pesquisa é baseada na Fonologia Entoacional Autossegmental e Métrica, que considera o acento tonal associado a uma sílaba tônica dentro de um constituinte prosódico, explicado pela Fonologia Prosódica. No português brasileiro, estudos prévios (Massini, 1991; Massini-Cagliari, 1992, 1993; Fernandes, 2007b; Tenani; Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2010; Toneli, 2014; Fernandes-Svartman; Romano, 2017; Toneli; Abaurre; Vigário, 2018) indicam que as proeminências tonais tendem a ocorrer, de maneira preferencial, em sílabas tônicas, constitutivas de palavras fonológicas.

A identificação de tonicidade nas palavras fonológicas dos enunciados sintéticos é executada a partir da inspeção auditiva das ocorrências e, em especial, da análise acústica,

[82] Em conformidade com o processo neurolinguístico de produção da fala, deve-se retomar a informação de que, segundo Massini-Cagliari e Cagliari (2012), a programação prosódica inicial de um enunciado envolve a extensão temporal, a entoação, o ritmo, a velocidade de fala e a acentuação, ou seja, princípios que orientam o desenvolvimento adequado do contorno melódico. Esses mesmos princípios fundamentam a priorização, na pesquisa, da transcrição dos eventos tonais atribuídos à primeira e à última sílaba tônica, bem como à fronteira direita da frase entoacional, de forma que seja possível a observação do delineamento do contorno melódico dos enunciados declarativos neutros.

sem a realização de testes de percepção com um grupo controlado de participantes. A entoação declarativa neutra é analisada a partir da interação entre a Fonética e a Fonologia, a fim de descrever os parâmetros acústicos da F_0 que, em conjunto com a estrutura fonológica, geram a configuração do contorno melódico (Soncin; Tenani, 2016; Soncin; Tenani; Berti, 2017, 2019). O trabalho segue os critérios utilizados na literatura sobre a prosódia do português brasileiro (Tenani, 2002; Fernandes, 2007b; Tenani; Fernandes-Svartman, 2008; Frota *et al.*, 2015; Frota; Moraes, 2016; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), que associam os acentos tonais às sílabas proeminentes de acordo com as variações acústicas da F_0. Nesse sentido, a abordagem metodológica do estudo é coerente com as pesquisas anteriores sobre a entoação do português brasileiro e assegura que a identificação dos acentos tonais seja realizada de forma objetiva e replicável, sem a necessidade de avaliações que dependem, em um primeiro momento, da percepção de um grupo controlado de falantes, apesar de tais verificações serem úteis para a ratificação dos resultados.

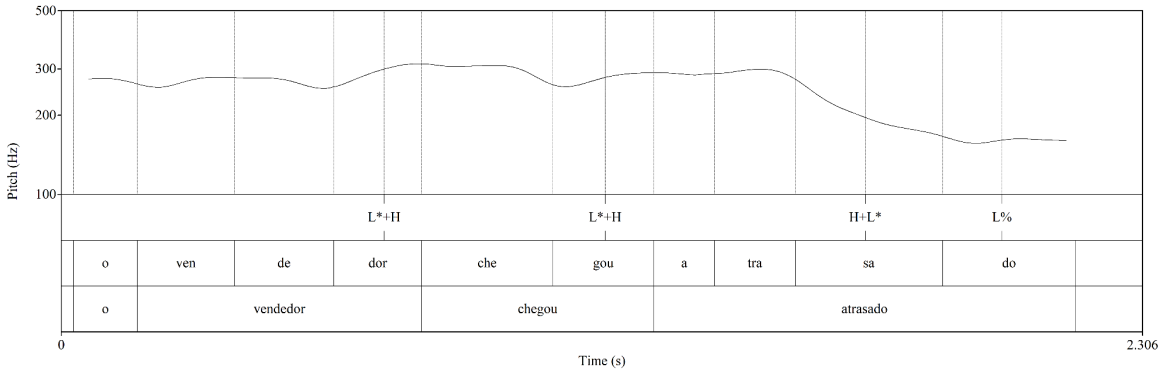
Com base na abordagem adotada para a entoação, apenas as variações fonéticas distintivas são integradas à representação fonológica dos contornos melódicos, ainda que outras também possam ter importância na descrição do sinal acústico (Soncin; Tenani, 2016). Nas Figuras 34 e 35, é exemplificada a distribuição tonal dos enunciados "O vendedor chegou atrasado" e "A pesquisadora terminou os trabalhos".[83] Para a transcrição entoacional dos enunciados com o auxílio do *software* Praat, são criadas as camadas de notação *tones* (tons), *syllables* (sílabas) e *words* (palavras), tal como feito nos estudos de Fernandes (2007a, 2007b), Tenani e Fernandes-Svartman (2008) e Soncin e Tenani (2016). A representação ilustrativa da F_0 desses enunciados é realizada com base nos quatro critérios explicados em parágrafos precedentes.[84]

---

[83] Nas figuras das seções "Metodologia" e "Resultados e Discussões", são adotadas, com as adaptações necessárias, as convenções da Fonologia Entoacional Autossegmental e Métrica e do sistema P-ToBI: L (*low*) indica um tom baixo; H (*high*) assinala um tom alto; o asterisco (*) marca o alvo tonal associado a uma proeminência melódica (acento tonal); o símbolo de adição (+) revela a ocorrência de dois tons em um mesmo evento tonal (acento tonal ou tom de fronteira); o símbolo de porcentagem (%) representa um tom de fronteira, que delimita uma frase entoacional; as linhas contínuas correspondem ao contorno entoacional obtido a partir da variação da F_0, com a suavização e a interpolação necessárias; e as barras verticais (| |) sinalizam a segmentação em palavras e sílabas ortográficas.

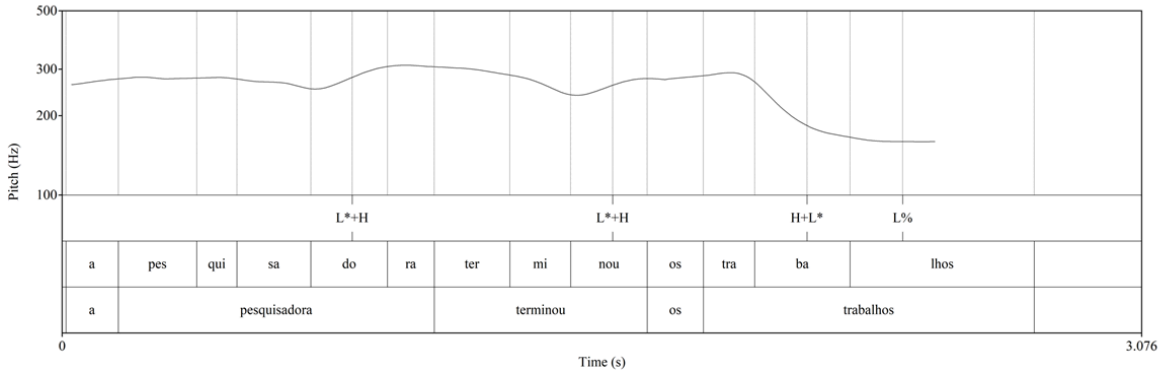
[84] A pesquisa emprega o sistema P-ToBI para a notação tonal, adaptado, conforme os objetivos do estudo, para a investigação das propriedades exclusivamente fonológicas do contorno entoacional. A adaptação prioriza a inclusão apenas das variações fonéticas relevantes e desconsidera, tanto quanto possível, as microalterações ou os detalhes acústicos sem quaisquer implicações para a organização gramatical da entoação declarativa neutra no português brasileiro. Nas figuras ilustrativas da distribuição tonal, são exibidas as camadas dos tons, das sílabas e das palavras, sem a presença do oscilograma, do espectrograma e dos índices numéricos de notação prosódica (*Break Indices*).

Figura 34 – Distribuição tonal do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-A, com suavização da F_0 e interpolação fonética



Fonte: Elaborado pelo autor (2025).

Figura 35 – Distribuição tonal do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Standard-A, com suavização da F_0 e interpolação fonética



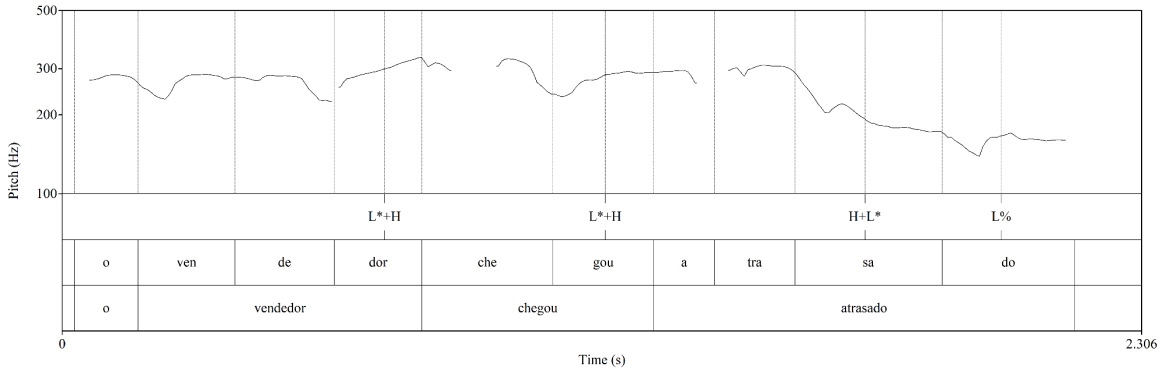
Fonte: Elaborado pelo autor (2025).

Em determinados estudos fonéticos sobre a prosódia da língua portuguesa, tais como os de Barbosa (1996, 1999, 2006), Lucente (2012) e Constantini (2014), as sílabas são segmentadas em unidades que se iniciam no início de uma vogal e se encerram no início da vogal seguinte, com a inclusão das consoantes entre elas (Constantini; Barbosa, 2015). Todavia, como o objetivo da pesquisa é interpretar os resultados gramaticalmente, utiliza-se o conceito de sílaba fonológica, conforme a análise de trabalhos como os de Camara Jr. (1969, 1975, 2015), Collischonn (2005), Massini-Cagliari (2015) e Cristófaro Silva (2022).[85]

[85] A abordagem dinâmica do ritmo, apresentada em trabalhos como os de Barbosa (1996, 1999, 2006), estabelece a unidade de vogal a vogal, relativa ao grupo *inter-perceptual-center*, como a base microrrítmica. Contudo, outros estudos acústicos sobre a fala natural do português brasileiro, como os de Massini (1991) e Massini-Cagliari (1992, 1993), indicam que, nas sílabas tônicas fonéticas, tanto a vogal quanto as consoantes adjacentes se alongam, o que pode comprometer a adequação de uma segmentação estrita de vogal a vogal para modelar a tonicidade primária e o pé métrico. Na pesquisa, observa-se, na fala sintética, o mesmo padrão descrito por Massini (1991) e Massini-Cagliari (1992, 1993) para a fala natural. Por esse motivo, opta-se pela concepção clássica de sílaba, originária de pesquisas com uma orientação gramatical, desde o Estruturalismo, pois ela preserva a configuração do pé métrico adotada, décadas depois, pela Fonologia Prosódica e reflete o padrão dos dados de fala sintética, em que o correlato acústico da tonicidade primária (aumento de duração)

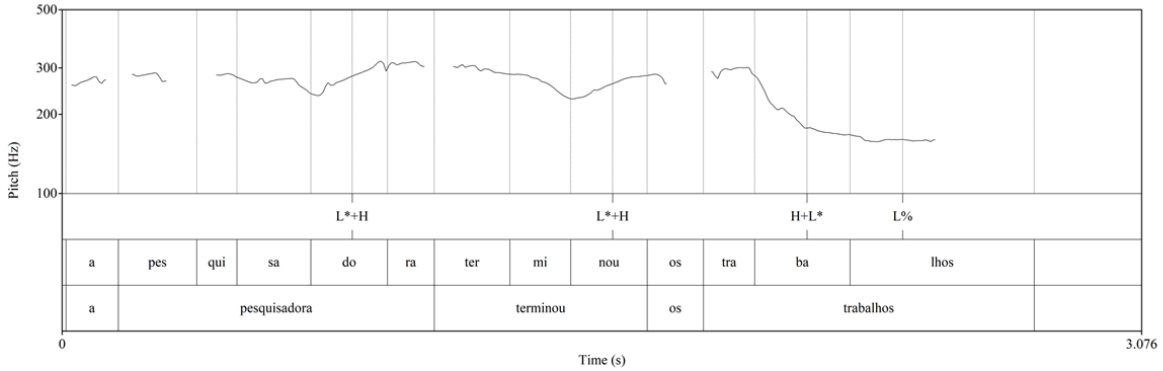
A título de exemplificação, nas Figuras 36 e 37, é ilustrada a distribuição tonal dos enunciados “O vendedor chegou atrasado” e “A pesquisadora terminou os trabalhos”, com a ausência do emprego dos critérios referentes à suavização da F_0 e à interpolação fonética. Entretanto, ao adotar os critérios mencionados, como nas Figuras 34 e 35, são verificadas, sem que haja interrupções no contorno da F_0, somente as curvas entoacionais capazes de desempenhar uma função linguística no sistema fonológico do português brasileiro, o que corrobora o emprego dos referidos recursos.

Figura 36 – Distribuição tonal do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-A, sem suavização da F_0 e interpolação fonética



Fonte: Elaborado pelo autor (2025).

Figura 37 – Distribuição tonal do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Standard-A, sem suavização da F_0 e interpolação fonética



Fonte: Elaborado pelo autor (2025).

O sistema de transcrição entoacional adotado esclarece que, a depender dos objetivos da transcrição, há a possibilidade de inclusão de detalhes fonéticos referentes a pico atrasado

também incide sobre os ataques e as codas adjacentes. Dessa forma, a segmentação de vogal a vogal não é utilizada, uma vez que conflita com a representação fonológica do pé métrico e não captura o alongamento segmental completo das sílabas tônicas fonéticas.

(<), pico adiantado (>) ou pico elevado (¡).[86] Como o propósito da pesquisa é uma interpretação fonológica do fenômeno, eventos fonéticos como os mencionados não são incluídos na transcrição entoacional dos enunciados declarativos neutros do português brasileiro. Dessa maneira, somente os eventos fonológicos, que desempenham uma função distintiva e um papel importante na organização do sistema linguístico, são considerados na transcrição entoacional das ocorrências.

A decisão teórico-metodológica se alinhada com o posicionamento de Córdula (2012, 2013), que, ao realizar uma descrição da entoação do português brasileiro e do inglês norte-americano com base na Fonologia Entoacional Autossegmental e Métrica, não considera, por exemplo, os casos de tons em degrau acima (*upstep*) ou degrau abaixo (*downstep*) e o alinhamento atrasado ou adiantado dos tons. Segundo a autora, com a qual a pesquisa concorda, essa conduta é justificada "pela não presença da descrição dessa tipologia de tom (*upstep/downstep*) na proposta precursora de Pierrehumbert (1980) e pela não função distintiva fonológica do alinhamento atrasado ou adiantado dos tons" (Córdula, 2013, p. 207, grifos da autora).

Embora o degrau acima e o degrau abaixo sejam discutidos na proposta de Pierrehumbert (1980), eles não se enquadram em uma tipologia específica de tons, como os acentos tonais e os tons de fronteira, pois envolvem a realização fonética dos eventos tonais. O mesmo se aplica aos casos de alinhamento atrasado ou antecipado dos tons, já que essas questões são exclusivamente acústicas, ao invés de fonológicas. Elas se relacionam às sincronizações de tempo do pico e do vale com os segmentos sonoros (Madureira, 2016), sem a constituição de um aspecto distintivo do sistema fonológico do português brasileiro, como o contraste entre os acentos tonais L*+H e H+L*, em que o primeiro indica uma ascendência melódica, enquanto o segundo configura uma descendência tonal.

A justificativa em discussão também é respaldada pelo trabalho de Truckenbrodt, Sandalo e Abaurre (2009). Ao investigarem a entoação do português brasileiro, os autores verificam que os enunciados declarativos enfáticos parecem ter o mesmo contorno entoacional nuclear H+L* L% dos enunciados declarativos neutros. Segundo os pesquisadores, uma distinção consistente entre as declarações enfáticas e as neutras reside na maior escala tonal das primeiras, um aspecto relacionado à gama de variação tonal (*pitch range*), por vezes chamada de amplitude tonal. Nesse sentido, os autores mantêm, no nível fonológico, a representação entoacional H+L* L% para o contorno entoacional nuclear tanto

[86] Informações disponíveis em: https://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/ToBI/ToBI_cv.html. Acesso em: 31 de janeiro de 2025.

das declarações neutras quanto das enfáticas e ressaltam que as enfáticas dispõem, no nível fonético, de uma F_0 mais alta em relação às neutras.

A pesquisa transcreve a estrutura entoacional dos dados de fala sintética somente com a representação de aspectos fonológicos, sem a inclusão de detalhes exclusivamente acústicos. No entanto, tal posicionamento não desconsidera a importância dos eventos fonéticos mencionados para a descrição entoacional. Esses eventos são relevantes e podem contribuir para a caracterização da expressividade da fala, uma vez que são capazes de gerar diferentes efeitos de sentido (Madureira, 2016).[87] Contudo, eles não são abordados na investigação, visto que a finalidade do estudo é mapear apenas as propriedades linguísticas da entoação declarativa neutra sintética e compará-las com a gramática do português brasileiro na fala natural.[88]

3.2.2 Determinação do padrão entoacional dos enunciados declarativos neutros

A segunda etapa é concernente à determinação do padrão entoacional do conjunto de enunciados declarativos neutros. A descrição dos tons atribuídos ao contorno melódico, realizada na etapa anterior, auxilia na determinação do padrão entoacional dos enunciados produzidos pela tecnologia de fala sintética do Google Cloud Text-to-Speech. Conforme detalhado na próxima seção, a estrutura entoacional predominante da declaração neutra na fala sintética é L*+H _____ H+L* L%.[89] No entanto, assim como ocorre na fala natural, há outra possível configuração entoacional na fala sintética, que é discutida e exemplificada na seção reservada à explanação dos resultados da pesquisa. Na Figura 38, é ilustrada a identificação dos eventos tonais do enunciado “Batata combina com peixe”, em que os tons L*+H, L*+H e H+L* são associados, nessa ordem, às sílabas tônicas “ta”, “bi” e

---

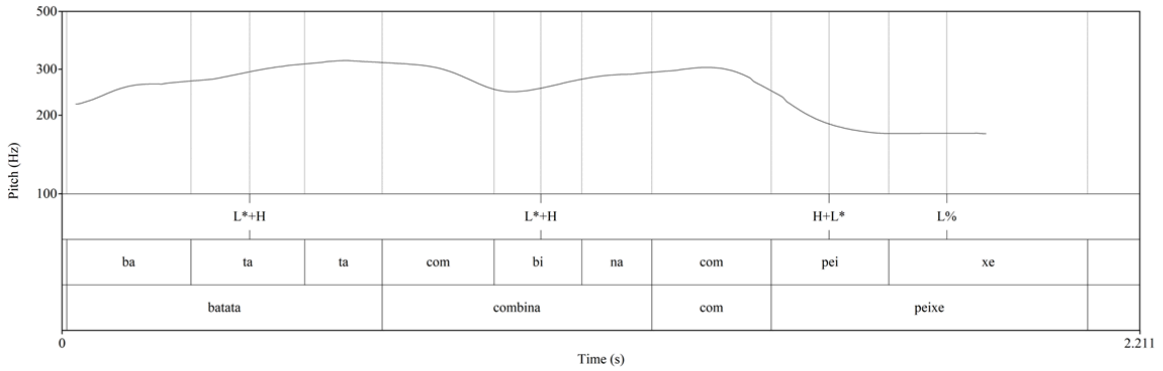
[87] Ainda que a fala sintética apresente, no nível fonológico, os contornos melódicos compatíveis com os da fala natural, podem ocorrer, no nível fonético, diferenças acústicas relacionadas à expressividade. Essas variações, entretanto, não se manifestam de maneira sistemática nem se estendem a todos os contextos e enunciados. Além disso, tais variações não devem ser compreendidas como limitações do sistema computacional, pois a própria fala natural também é caracterizada por modificações na realização acústica da entoação. Essa variabilidade decorre de fatores comunicativos, estilísticos e idiossincráticos que podem influenciar aspectos como o posicionamento do pico tonal ou a amplitude da variação tonal. Assim, é importante esclarecer que a intenção da pesquisa não é analisar a expressividade prosódica, mas descrever os padrões fonológicos da estrutura entoacional de enunciados declarativos neutros e compará-los com os da fala natural.

[88] Ressalta-se a existência de outras propostas para a descrição e análise da entoação, como as de Pike (1945), Abercrombie (1967), Halliday (1970), Bolinger (1986), Fónagy (1993) e Lucente (2008, 2012, 2017). No entanto, esse estudo se distancia dos trabalhos mencionados ao assumir um ponto de vista que considera uma relação entre a Fonologia Entoacional Autossegmental e Métrica e a Fonologia Prosódica, conforme tratado, por exemplo, em Frota e Vigário (2000), Frota (2000), Tenani (2002), Fernandes (2007b) e Serra (2009).

[89] A pesquisa é inspirada em trabalhos como os de Cunha (2000), Moraes (2008) e Silvestre e Cunha (2013) para a representação fonológica L*+H _____ H+L* L%.

“pei”, e o tom L% é associado à sílaba postônica “xe”.

Figura 38 – Eventos tonais do enunciado “Batata combina com peixe”, gerado pela voz pt-BR-Standard-A



Fonte: Elaborado pelo autor (2025).

3.2.3 Comparação dos aspectos entoacionais da fala sintética e da fala natural

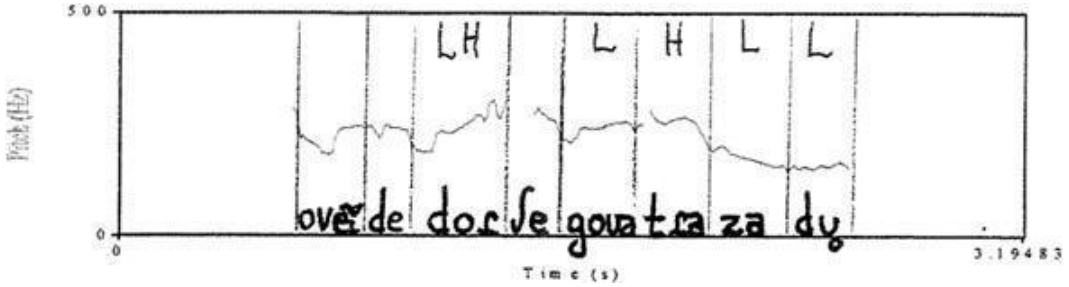
A terceira etapa consiste na comparação dos aspectos entoacionais dos enunciados sintéticos com os resultados obtidos por Tenani (2002) em uma investigação relativa à estrutura entoacional de enunciados declarativos neutros produzidos, com o auxílio de métodos experimentais, por falantes brasileiros. Justifica-se a escolha do estudo pioneiro de Tenani (2002) porque a autora também adota uma visão integrada entre a Fonologia Entoacional Autossegmental e Métrica e a Fonologia Prosódica para caracterizar a estrutura entoacional da declaração neutra no português brasileiro. Além disso, a pesquisa utiliza o *corpus* elaborado pela própria autora para a investigação desse parâmetro melódico.[90]

Como é possível observar, de modo pormenorizado, na próxima seção, a configuração definida por uma ascendência tonal no começo da sentença e por uma descendência melódica no término do enunciado corresponde ao padrão entoacional que predomina no conjunto de dados de fala sintética descritos, em concordância com o estudo de Tenani (2002) e os demais trabalhos prosódicos utilizados na pesquisa (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012;

[90] Para a discussão dos dados de fala sintética, além do estudo de Tenani (2002), consideram-se outros trabalhos sobre a entoação do português brasileiro (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b), a fim de fornecer uma base mais ampla e contextualizada para a interpretação dos resultados.

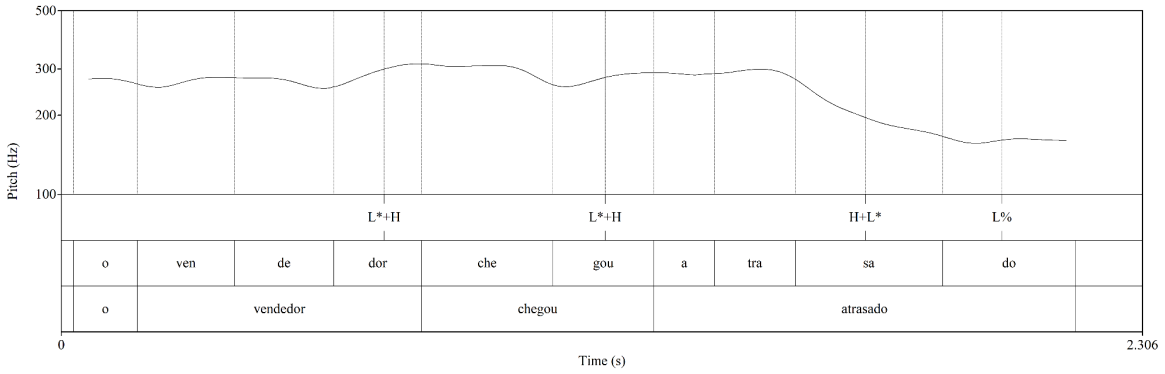
Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). Nas Figuras 39 e 40, é ilustrada a identificação dos eventos tonais do enunciado "O vendedor chegou atrasado", produzido por uma voz natural e por uma voz sintética. Nota-se que as duas ocorrências desse enunciado têm, no domínio fonológico, o mesmo contorno entoacional nuclear descendente, além do acento tonal ascendente associado ao começo das frases entoacionais (contorno entoacional pré-nuclear).

Figura 39 – Eventos tonais do enunciado "O vendedor chegou atrasado", produzido por uma voz natural



Fonte: Tenani (2002, p. 44).

Figura 40 – Eventos tonais do enunciado "O vendedor chegou atrasado", gerado pela voz pt-BR-Neural2-A



Fonte: Elaborado pelo autor (2025).

Ao adotar a comparação com os resultados advindos do *corpus* de Tenani (2002) e a etiquetagem tonal segundo a tradição do sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015), o estudo propõe um protocolo de análise entoacional que pode ser replicado e útil para a avaliação melódica da declaração neutra do português brasileiro em outros sistemas de síntese de fala. Esse protocolo, que combina a inspeção auditiva, a análise acústica e a representação fonológica, é uma contribuição metodológica, visto que estabelece um

procedimento padronizado para a verificação da correspondência entre o modelo computacional e o sistema gramatical, o que é benéfico tanto para os linguistas quanto para os cientistas da computação que desejam mensurar a fidelidade da entoação declarativa neutra sintética do português brasileiro em relação à natural.

Após a execução das etapas mencionadas, os resultados são analisados conforme uma abordagem quali-quantitativa. Busca-se, assim, obter uma interpretação robusta e consistente das características entoacionais dos enunciados examinados.

3.3 PESQUISA QUALI-QUANTITATIVA

A pesquisa quali-quantitativa, também conhecida como mista ou quali-quanti, integra métodos qualitativos e quantitativos na coleta de dados, com o propósito de proporcionar uma compreensão mais abrangente do fenômeno investigado (Paiva, 2019). Essa abordagem permite examinar a entoação declarativa neutra produzida pelo Google Cloud Text-to-Speech, por meio da análise qualitativa das características melódicas do português brasileiro e da validação quantitativa dos padrões fonológicos identificados.

No âmbito qualitativo, o estudo emprega métodos fonéticos para a identificação das propriedades prosódicas responsáveis pela estrutura entoacional do português brasileiro e recorre a princípios fonológicos que esclarecem a integração da prosódia ao sistema linguístico. Objetiva-se realizar uma análise detalhada de enunciados declarativos neutros produzidos, no português brasileiro, por mecanismos de síntese de fala da empresa Google. Essa perspectiva tem a finalidade de verificar os padrões entoacionais recorrentes, compará-los com a fala natural e refletir sobre o modo como o Google Cloud Text-to-Speech reproduz a estrutura melódica prevista, para as declarações neutras, pela prosódia do português brasileiro.

A orientação da pesquisa considera que a Fonética não corresponde somente à análise física da fala, assim como a Fonologia não se esgota na abstração de categorias formais. Ambas podem ser entendidas como duas disciplinas complementares, que lidam com as informações linguísticas codificadas em níveis distintos. Dessa forma, a pesquisa assume o caráter teórico-metodológico de investigar a entoação declarativa neutra como um fenômeno gramatical resultante da interface entre a representação fonológica e a manifestação fonética.

A menção aos métodos quantitativos, por sua vez, refere-se, na pesquisa, ao uso da estatística descritiva básica para a sistematização do material coletado, sem a realização de testes estatísticos inferenciais. A quantificação é realizada por meio da contagem dos eventos

tonais identificados nos enunciados sintéticos, o que permite calcular a frequência de cada padrão entoacional e a distribuição dos tons gramaticais em relação ao total de enunciados declarativos neutros analisados. Esses dados se organizam com o propósito de identificar tendências nos contornos entoacionais e realizar comparações entre os enunciados declarativos neutros sintéticos e os padrões descritos na literatura sobre a fala natural do português brasileiro.

Um dos motivos para a não utilização de testes estatísticos inferenciais é o número de ocorrências no *corpus* da pesquisa. Ademais, conforme observado na próxima seção, os enunciados declarativos neutros sintéticos apresentam características entoacionais com uma organização próxima à categórica. Esse aspecto torna desnecessária a aplicação de análises inferenciais, uma vez que os procedimentos estatísticos descritivos, alinhados à análise qualitativa, são suficientes para apresentar ao leitor os padrões entoacionais dos dados de fala sintética produzidos pelo Google Cloud Text-to-Speech.

Com a descrição do material e dos métodos, passa-se à apresentação dos resultados e das discussões da pesquisa, cuja prioridade é uma interpretação qualitativa do fenômeno.[91]

[91] Opta-se por apresentar os resultados e as discussões em uma seção integrada, pois, em estudos entoacionais, a interpretação das categorias gramaticais decorre da análise do sinal acústico correspondente às figuras das trajetórias da F_0 e dos eventos tonais. Para o pesquisador, a separação pode dificultar a leitura conjunta entre a figura e a interpretação. Além disso, a integração evita a ocorrência de redundâncias e assegura uma exposição imediata das evidências aos argumentos.

4 RESULTADOS E DISCUSSÕES

Esta seção, reservada à explanação dos resultados e das discussões, é dividida em três partes. A primeira subseção é relacionada aos acentos tonais, a segunda, ao tom de fronteira, e a terceira, às configurações entoacionais do conjunto de enunciados declarativos neutros gerados por um sistema de síntese de fala baseado em IA.

A descrição e a análise dos dados são pautadas nos pressupostos teórico-metodológicos da Fonética Acústica, da Fonologia Entoacional Autossegmental e Métrica e da Fonologia Prosódica, a fim de que haja a verificação dos aspectos físicos e das propriedades fonológicas da entoação declarativa neutra. Em outras palavras, além de caracterizada a variação fonética da F_0, é fornecida uma interpretação fonológica à realização acústica da entoação declarativa neutra.[92]

A caracterização da variação fonética da F_0 se concentra apenas na descrição da configuração melódica dos pontos relevantes da estrutura tonal, sem a realização de cálculos avançados de medidas acústicas. Deve-se salientar que, na pesquisa, a análise acústica é utilizada como uma ferramenta para a compreensão das propriedades fonológicas da entoação sintética. Ao longo da seção, é feita uma comparação dos aspectos entoacionais dos enunciados sintéticos com os resultados obtidos por Tenani (2002) em uma investigação relativa à estrutura entoacional de enunciados declarativos neutros produzidos, com o auxílio de métodos experimentais, por falantes brasileiros. Ademais, em ocasiões oportunas, é estabelecido um diálogo com outros estudos acerca das características fonético-fonológicas da entoação do português brasileiro.[93]

A seguir, são detalhados os acentos tonais da frase entoacional que constitui a declaração neutra.

4.1 ACENTOS TONAIS

Conforme discutido na seção de fundamentação teórica, os acentos tonais são associados às proeminências da frase entoacional, e o domínio prosódico de atribuição dos

[92] O estudo não se limita à descrição das semelhanças superficiais da naturalidade. Na verdade, ele busca demonstrar se as categorias entoacionais são recuperáveis e sistemáticas nas saídas acústicas de modernos modelos de síntese de fala. Trata-se de uma conduta que posiciona a investigação em um debate epistemológico que questiona se as representações fonológicas são propriedades intrinsecamente humanas ou princípios estruturais passíveis de serem modelados estatisticamente por um sistema computacional.

[93] A falta de experimentos psicoacústicos com a participação de ouvintes humanos é decorrente do tempo reduzido destinado à pesquisa. Mesmo assim, a análise acústico-fonológica, aliada à literatura especializada, garante a validade dos resultados.

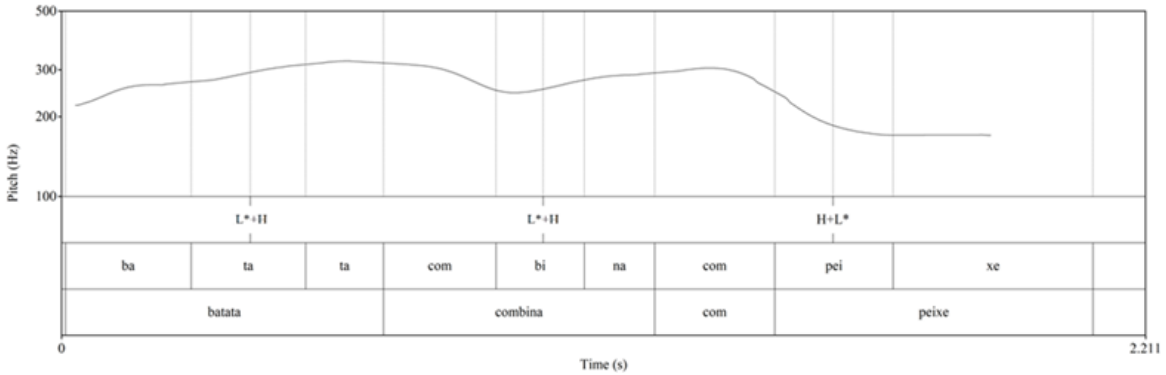
acentos tonais, na gramática prosódica do português brasileiro, é a palavra fonológica, em que há somente uma sílaba tônica, caracterizada, de um ponto de vista fonético, por um aumento de duração, em comparação com as sílabas átonas (Massini-Cagliari, 1993). Quanto ao acento de palavra fonológica, soma-se à discussão que as sílabas postônicas tendem a ser marcadas, no nível fonético, por um decréscimo de intensidade (Massini-Cagliari, 2019), reconhecida pela variação de volume (fraco ou forte) e definida como a quantidade de energia de uma onda sonora (Cristófaro Silva, 2015). Ademais, a tonicidade, no enunciado de uma língua de ritmo acentual, pode ser reforçada ou diminuída por meio da variação entoacional da fala (Cagliari, 2012b).[94] À vista disso, identifica-se, com o suporte do *software* Praat, se há a associação de acentos tonais às palavras fonológicas que compõem as frases entoacionais do *corpus*.

A atribuição de acento tonal ocorre nas 189 palavras fonológicas da amostra examinada. O resultado em questão, apresentado em Tojeira-Ramos e Massini-Cagliari (no prelo), reforça a alta densidade tonal do português brasileiro, atestada por Frota *et al.* (2015), Frota e Moraes (2016) e Fernandes-Svartman (2024a, 2024b) em dados naturais, e corrobora o fato de a palavra fonológica ser, nessa variedade linguística, o domínio prosódico relevante para a associação de acentos tonais (Massini, 1991; Massini-Cagliari, 1992, 1993; Fernandes, 2007b; Tenani; Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2010; Toneli, 2014; Fernandes-Svartman; Romano, 2017; Toneli; Abaurre; Vigário, 2018). As Figuras de 41 a 47 ilustram exemplos em que há a atribuição de acentos tonais às palavras fonológicas do conjunto de enunciados descritos.[95]

[94] Apesar de ser classificado, por certos autores, como uma variedade linguística caracterizada por um ritmo acentual (Cagliari, 1981, 1984, 1985; Moraes; Leite, 1992; Massini, 1991; Massini-Cagliari, 1992, 1995), fato com o qual a pesquisa concorda, há tanto processos fonológicos quanto velocidades e estilos de fala que podem contribuir para a ocorrência de um ritmo silábico no português brasileiro (Abaurre-Gnerre, 1981; Cagliari; Abaurre, 1986).

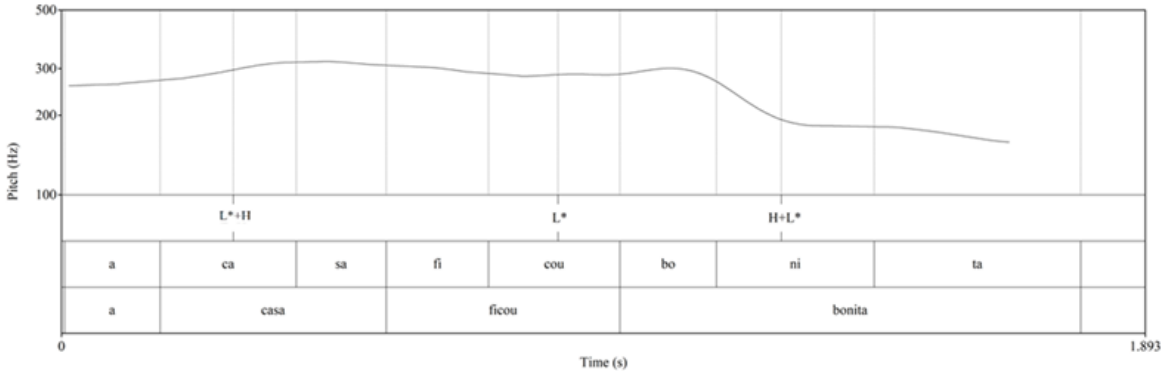
[95] As 189 palavras fonológicas analisadas são "cabeça" de frase fonológica. Como informado na seção anterior, as frases entoacionais do *corpus* são constituídas de três frases fonológicas não ramificadas, cada qual contendo uma palavra fonológica (Tenani, 2002). Sugere-se, para pesquisas futuras, a análise de sentenças com frases fonológicas formadas por mais de uma palavra fonológica. Dessa maneira, é possível verificar a frequência de atribuição de acento tonal às palavras fonológicas não "cabeça" de frase fonológica.

Figura 41 – Acentos tonais associados às palavras fonológicas do enunciado “Batata combina com peixe”, gerado pela voz pt-BR-Standard-A



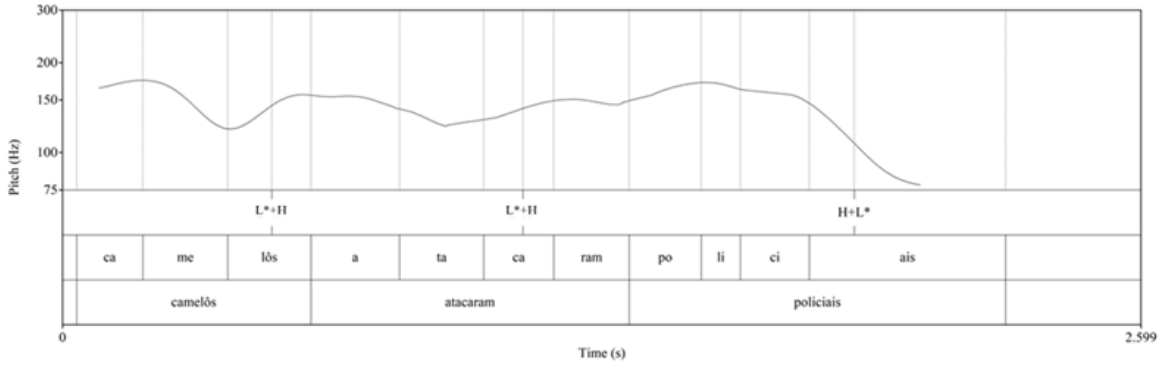
Fonte: Elaborado pelo autor (2025).

Figura 42 – Acentos tonais associados às palavras fonológicas do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Standard-A



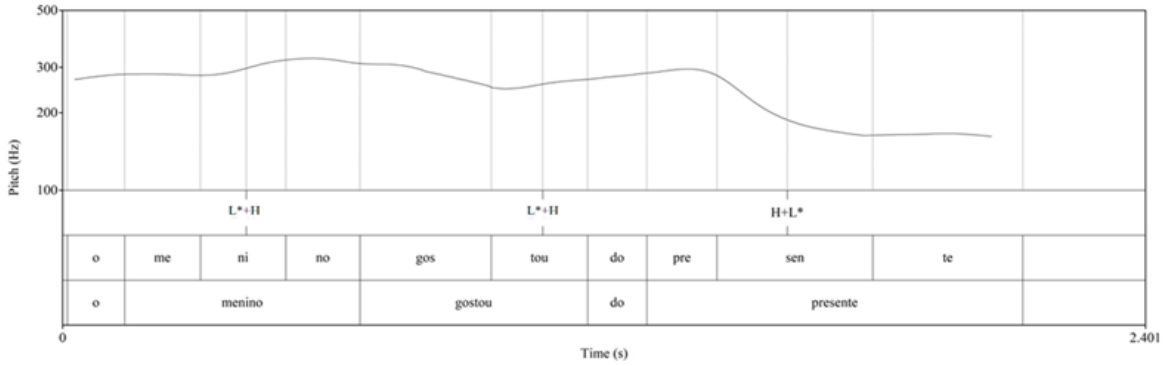
Fonte: Elaborado pelo autor (2025).

Figura 43 – Acentos tonais associados às palavras fonológicas do enunciado “Camelôs atacaram policiais”, gerado pela voz pt-BR-Standard-B



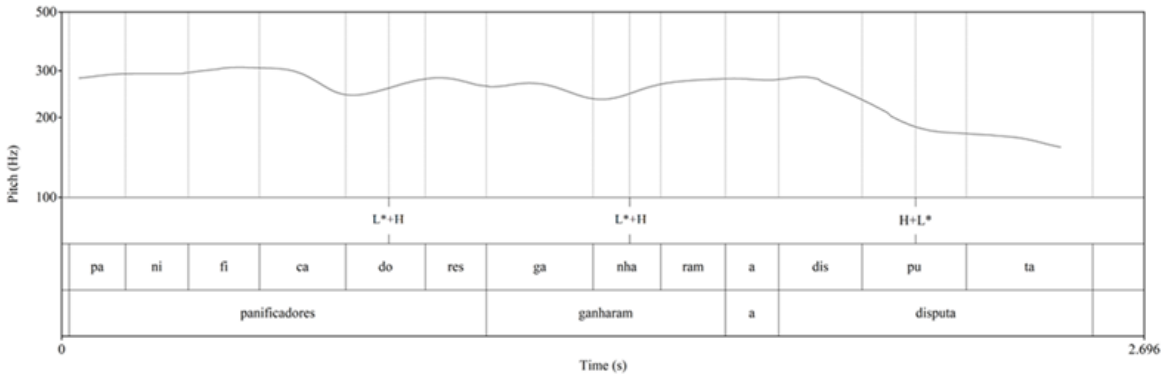
Fonte: Elaborado pelo autor (2025).

Figura 44 – Acentos tonais associados às palavras fonológicas do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Standard-A



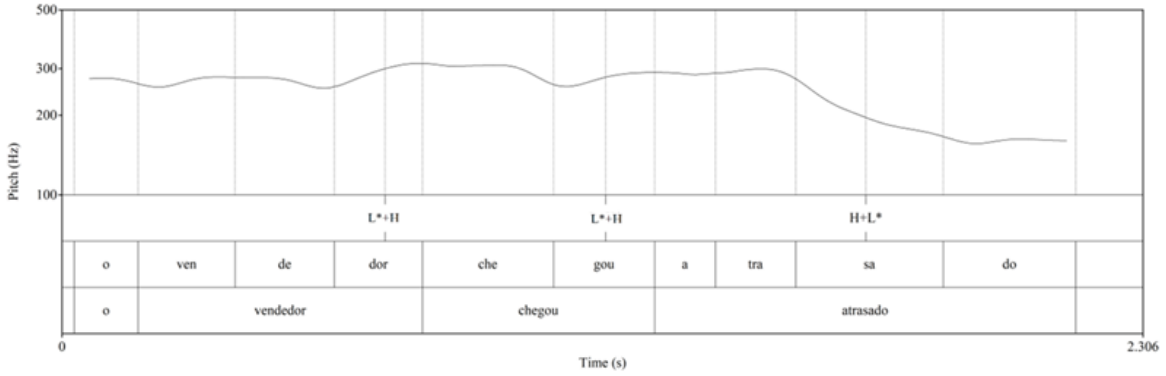
Fonte: Elaborado pelo autor (2025).

Figura 45 – Acentos tonais associados às palavras fonológicas do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Standard-A



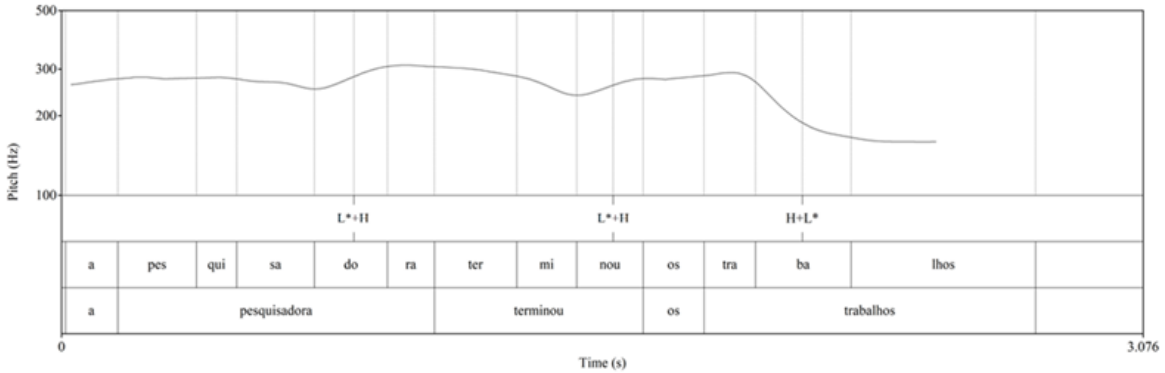
Fonte: Elaborado pelo autor (2025).

Figura 46 – Acentos tonais associados às palavras fonológicas do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-A



Fonte: Elaborado pelo autor (2025).

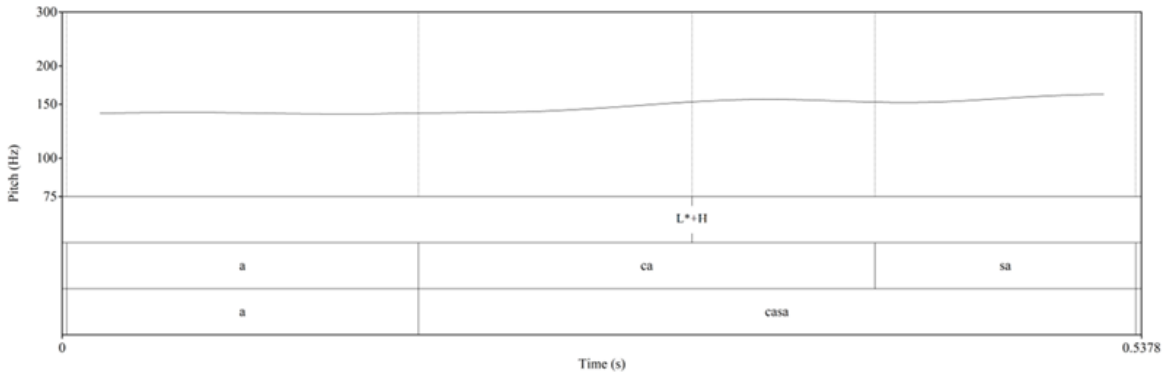
Figura 47 – Acentos tonais associados às palavras fonológicas do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Standard-A



Fonte: Elaborado pelo autor (2025).

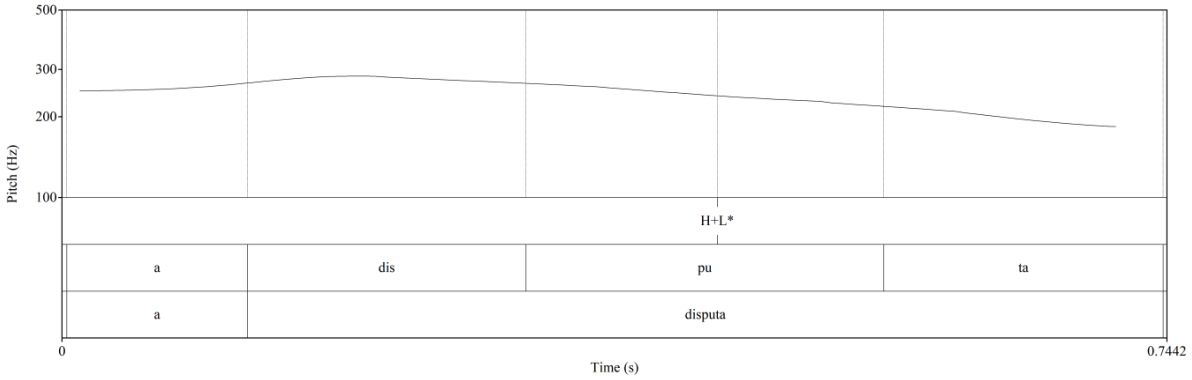
Segundo Tojeira-Ramos e Massini-Cagliari (no prelo), em uma parcela dos dados analisados, não ocorre o estabelecimento de uma relação de isomorfismo entre os componentes fonológico e morfológico da gramática, já que, em estruturas como “a casa” e “a disputa”, há, para cada forma, somente uma palavra fonológica (/aˈkaza/ e /adiSˈputa/), fonemicamente transcrita em concordância com Camara Jr. (1969, 1975, 2015), e duas palavras morfológicas (“a” e “casa”, e “a” e “disputa”). Nas Figuras 48 e 49, são exemplificadas ocorrências em que os acentos tonais incidem apenas sobre as sílabas tônicas /ˈka/ e /ˈpu/ das palavras fonológicas. Em “a casa” e “a disputa”, o elemento “a” (monossílabo átono), classificado como uma palavra morfológica com o estatuto de artigo, não recebe a atribuição de acento primário e tonal, visto que é definido como um marcador pré-nominal, sem tonicidade, associado ao substantivo (Castilho, 2019). Desse modo, a palavra fonológica (e não a morfológica) é, nos dados de fala sintética, o constituinte prosódico a ser considerado para a atribuição de acentos tonais (Tojeira-Ramos; Massini-Cagliari, no prelo), assim como ocorre na fala natural, com base nos trabalhos mencionados acima.

Figura 48 – Acento tonal associado à palavra fonológica “a casa”, extraída do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Wavenet-B



Fonte: Elaborado pelo autor (2025).

Figura 49 – Acento tonal associado à palavra fonológica “a disputa”, extraída do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Standard-C



Fonte: Elaborado pelo autor (2025).

Nas ocorrências de fala sintética estudadas, cada sentença constitui, na hierarquia prosódica, uma frase entoacional (coextensiva a um enunciado fonológico), como nos dados de fala natural descritos por Tenani (2002). A observação de pistas fonéticas auxilia na comprovação de que as sentenças examinadas formam somente uma frase entoacional. Por exemplo, nas ocorrências analisadas, há a presença de uma pausa silenciosa apenas para delimitar a fronteira entoacional, sem que haja uma reestruturação do domínio prosódico em outros constituintes entoacionais.[96] Além disso, o término das sentenças é marcado por uma

[96] Concorda-se com Soncin, Tenani e Berti (2017) quanto ao questionamento da natureza da pausa sob a ótica da percepção. As autoras defendem que a pausa ultrapassa a condição de um simples momento de silêncio e depende da identificação da fronteira entoacional no plano fonológico, sinalizada pela variação da F_0 no plano fonético. A pausa se situa em um contexto linguístico definidor, em que as informações simbólicas atuam simultaneamente na percepção dos fenômenos fonético-acústicos. Por meio de um teste de percepção, as autoras verificam a influência de diferentes tipos de informação após o controle das variáveis na preparação dos estímulos auditivos. Os resultados indicam que, para a percepção da pausa, não basta o padrão acústico característico do silêncio, pois também há o envolvimento de informações simbólicas relacionadas à representação fonológica da entoação, constituída a partir da interação com os elementos sintático-semânticos da língua analisada.

evidente descendência melódica, seguida, no movimento tonal de fronteira, de um *reset* (Serra, 2009). Algumas discussões sobre essas e outras pistas usadas para identificar as frases entoacionais podem ser encontradas, por exemplo, em Frota e Vigário (2000), Tenani (2002), Fernandes (2007b), Frota *et al.* (2007), Serra (2009), Soncin (2017), Soncin, Tenani e Berti (2017, 2019) e Fernandes-Svartman (2024a, 2024b).

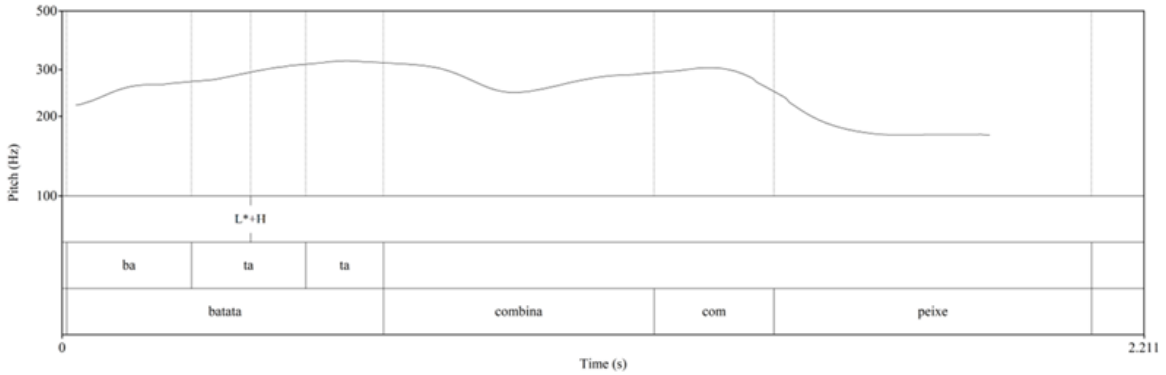
Com detalhes e exemplificações, nas subseções a seguir, é descrito e analisado, em um primeiro momento, o acento tonal atribuído à primeira sílaba tônica da frase entoacional. Já em um segundo momento, é descrito e analisado o acento tonal atribuído à última sílaba tônica da frase entoacional.

4.1.1 Acento tonal inicial

No conjunto de enunciados declarativos neutros, o acento tonal atribuído à primeira sílaba tônica é L*+H. Esse acento tonal ocorre, sem exceção, nos 63 dados da amostra. No plano fonético, ele é caracterizado por uma ascendência melódica, decorrente de um aumento da variação da F_0 que a torna, em termos auditivos, mais aguda (Cagliari; Massini-Cagliari, 2003).[97] Um som agudo é caracterizado por apresentar uma maior concentração de energia nas frequências altas (Maia, 1985). As Figuras de 50 a 56 ilustram ocorrências em que o acento tonal ascendente é atribuído às sílabas tônicas iniciais dos enunciados descritos. Opta-se por exibir, nas figuras, apenas a transcrição dos eventos tonais relevantes para a discussão, assim como Santos e Fernandes-Svartman (2020).

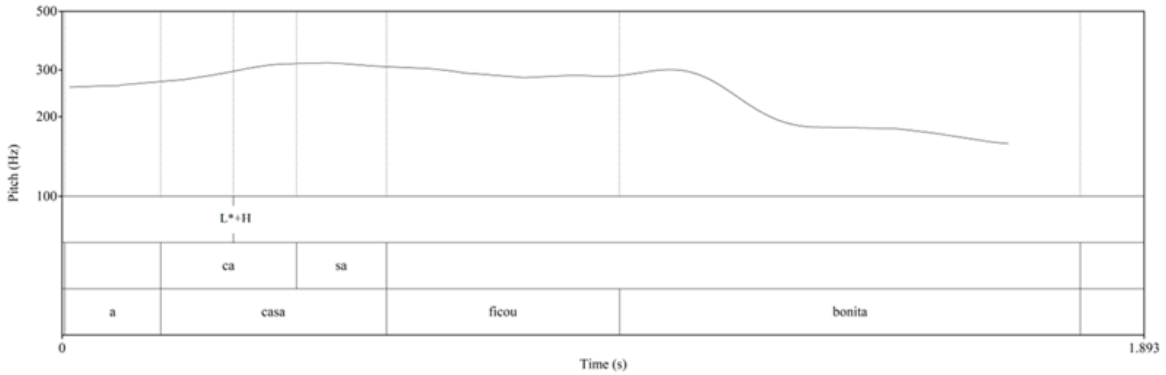
[97] Como é possível observar nas figuras ilustrativas da análise entoacional dos dados de fala sintética, há casos em que a ascendência melódica no início dos enunciados apresenta uma menor saliência em comparação com a verificada em outras figuras. Trata-se de um fato de natureza exclusivamente fonética, e não fonológica, que não interfere nos resultados da pesquisa. A interpretação do trabalho é fonológica, com o objetivo de examinar as propriedades gramaticais associadas à entoação declarativa neutra. A configuração fonética da ascendência melódica varia de acordo com diferentes fatores técnicos, relacionados ao tipo de voz analisada, e sonoros, vinculados à tonalidade e à qualidade da vogal tônica e das consoantes adjacentes, à extensão temporal da palavra fonológica e ao número de sílabas átonas pretônicas e postônicas, por exemplo. Tal variação não representa uma irregularidade metodológica ou analítica, visto que a melodia incidente nessa porção do enunciado não corresponde ao núcleo entoacional, ou seja, à região de maior proeminência do contorno melódico.

Figura 50 – Acento tonal associado à primeira sílaba tônica do enunciado “Batata combina com peixe”, gerado pela voz pt-BR-Standard-A



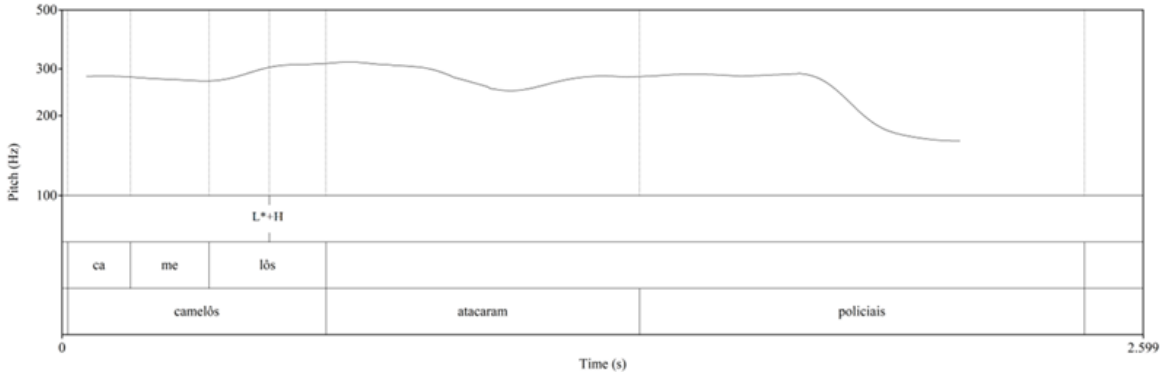
Fonte: Elaborado pelo autor (2025).

Figura 51 – Acento tonal associado à primeira sílaba tônica do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Standard-A



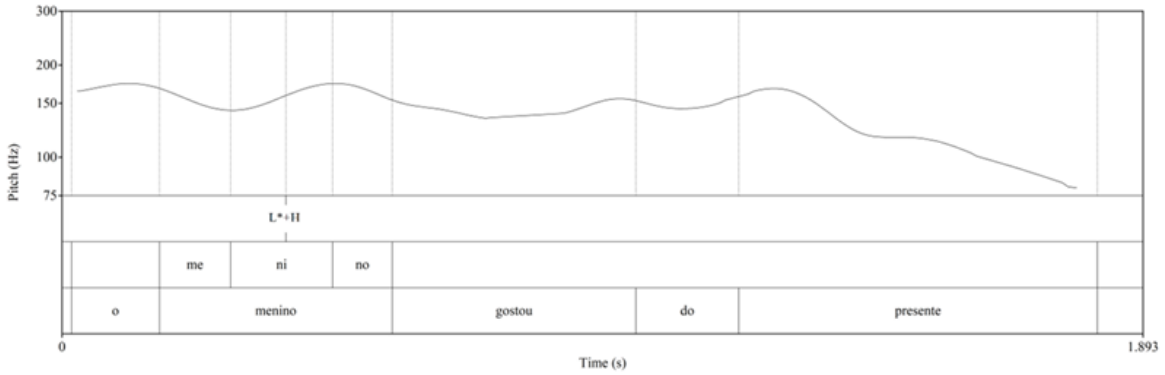
Fonte: Elaborado pelo autor (2025).

Figura 52 – Acento tonal associado à primeira sílaba tônica do enunciado “Camelôs atacaram policiais”, gerado pela voz pt-BR-Standard-A



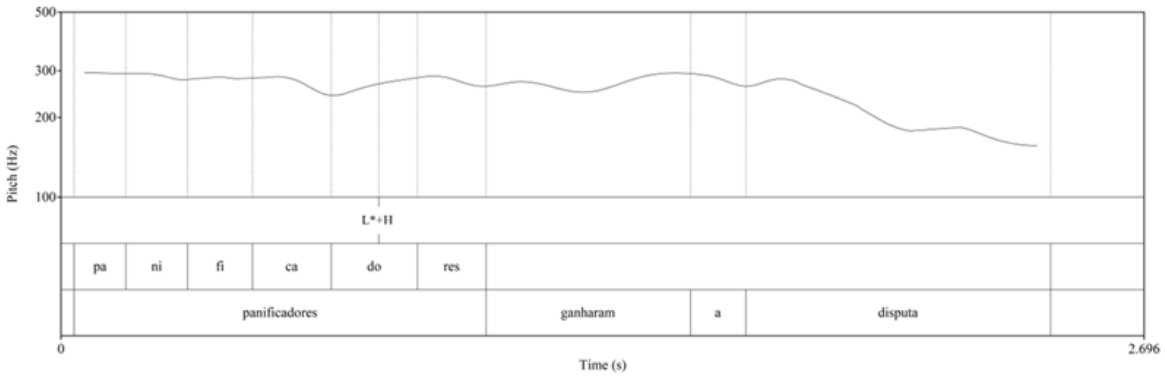
Fonte: Elaborado pelo autor (2025).

Figura 53 – Acento tonal associado à primeira sílaba tônica do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Wavenet-B



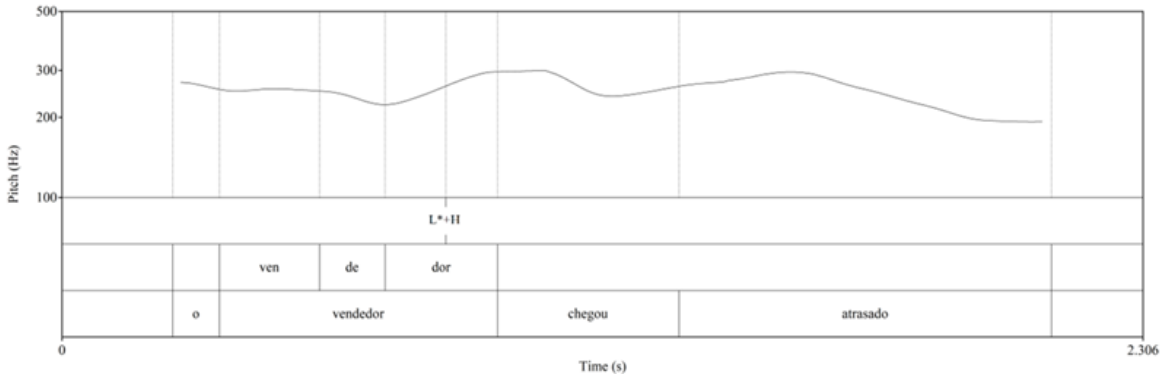
Fonte: Elaborado pelo autor (2025).

Figura 54 – Acento tonal associado à primeira sílaba tônica do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Neural2-A



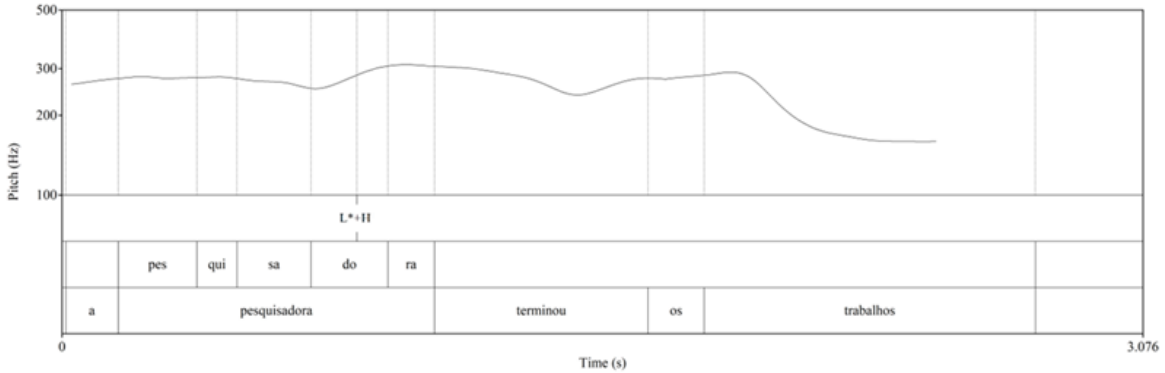
Fonte: Elaborado pelo autor (2025).

Figura 55 – Acento tonal associado à primeira sílaba tônica do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-C



Fonte: Elaborado pelo autor (2025).

Figura 56 – Acento tonal associado à primeira sílaba tônica do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Standard-A



Fonte: Elaborado pelo autor (2025).

A pesquisa indica que o acento tonal ascendente corresponde a um padrão comum no começo de enunciados declarativos neutros. A análise quantitativa confirma que a ocorrência regular do acento tonal ascendente é estatisticamente significativa e não decorre de um processo aleatório. Dessa maneira, torna-se possível prever, com confiabilidade, a presença do acento tonal ascendente no início de enunciados declarativos neutros quando se sintetiza, por meio do Google Cloud Text-to-Speech, a fala do português brasileiro. Essas descobertas sugerem que a síntese de fala do produto investigado tende a seguir esse padrão melódico ao gerar os enunciados declarativos neutros.

Em relação à gramática do português brasileiro, Tenani (2002) observa que, em enunciados declarativos neutros, há um acento tonal ascendente associado à primeira sílaba tônica da frase entoacional. Esse padrão melódico, caracterizado por uma elevação inicial na curva da F_0, corresponde ao contorno entoacional pré-nuclear e é representado, na pesquisa atual, pela notação fonológica L*+H, que indica um tom baixo na sílaba tônica, seguido de uma ascendência na sílaba postônica. A configuração observada confirma os resultados de outros estudos que descrevem esse padrão ascendente no início dos enunciados declarativos do português brasileiro (Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Frota *et al.*, 2015; Castelo, 2016; Fernandes-Svartman; Romano, 2017; Fernandes-Svartman, 2024a). Além disso, a análise dos enunciados gerados pela tecnologia de síntese de fala do Google Cloud Text-to-Speech evidencia a produtividade desse contorno melódico e sinaliza uma correspondência com os dados da fala natural. Esses resultados fortalecem a adoção da notação fonológica L*+H como uma possível representação do padrão ascendente no contorno entoacional pré-nuclear dos enunciados declarativos neutros no português brasileiro.

Ao destacar a autonomia da Língua-I como um sistema computacional, Kato (1997)

proporciona um embasamento teórico para a interpretação dos resultados referentes ao acento tonal inicial. A pesquisa salienta que a ascendência tonal no começo dos enunciados declarativos neutros sintéticos corresponde às regularidades melódicas descritas na fala natural. De acordo com a referida correspondência, os algoritmos de IA geram os padrões tonais compatíveis com a prosódia português brasileiro. Sob essa perspectiva, a fala sintética resulta de padrões estatísticos que reproduzem as regularidades simbólicas associadas à base do conhecimento linguístico dos falantes (Chomsky, 1993).

À luz das discussões precedentes, pode-se observar que, no tocante aos acentos tonais associados à primeira sílaba tônica dos enunciados declarativos neutros, há semelhanças entre a fala sintética e a natural. Nos dois tipos de fala, é identificada a associação do acento tonal ascendente à proeminência inicial da declaração neutra.

Como não costumam desempenhar um papel gramatical distintivo na língua portuguesa, os acentos tonais associados à sílaba tônica medial dos enunciados declarativos neutros não são descritos. Entretanto, é importante afirmar que, nas sílabas tônicas mediais analisadas, há a atribuição de acentos tonais, com a preferência pela ascendência melódica. Trata-se de um resultado que corrobora a afirmação de Tenani (2002) de que, na entoação declarativa neutra do português brasileiro, há a tendência em atribuir tons às frases fonológicas intermediárias, além de evidenciar a preferência dessa variedade linguística pela alternância L H L H entre os tons, como discutido pela autora referenciada.

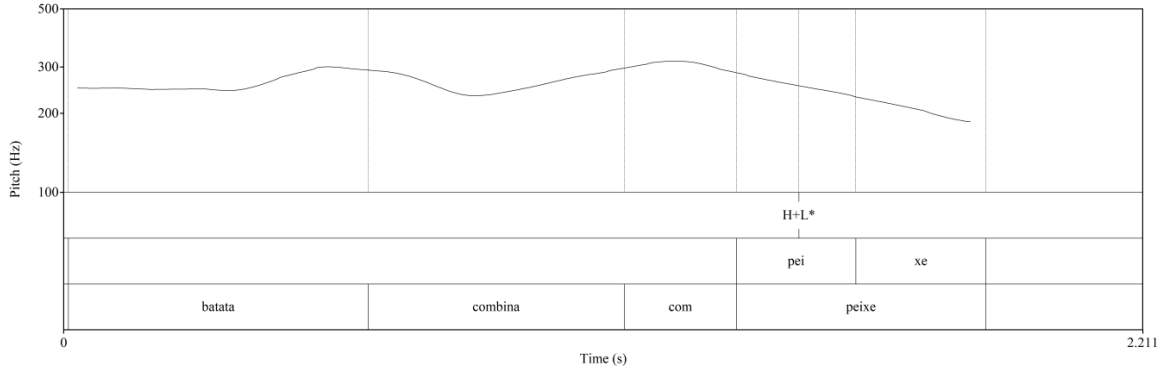
Na próxima subseção, é descrito e analisado o acento tonal atribuído à última sílaba tônica do contorno de entoação dos enunciados declarativos neutros.

4.1.2 Acento tonal final

No conjunto de enunciados declarativos neutros, o acento tonal atribuído à última sílaba tônica é H+L*. Esse acento tonal também ocorre, sem exceção, nos 63 dados da amostra. Na representação fonética, ele é expresso por um abaixamento ou descendência da curva melódica, resultante de uma diminuição da variação da F_0 que a torna, na percepção auditiva, mais grave (Cagliari; Massini-Cagliari, 2003). Um som grave é caracterizado por apresentar uma maior concentração de energia nas frequências baixas (Maia, 1985). A análise estatística comprova que o referido acento tonal se manifesta de forma regular, independentemente de variações aleatórias. Os resultados também indicam que o acento tonal final H+L* é uma característica comum de enunciados declarativos neutros sintéticos, com uma alta probabilidade de reprodução em situações futuras do Google Cloud Text-to-Speech.

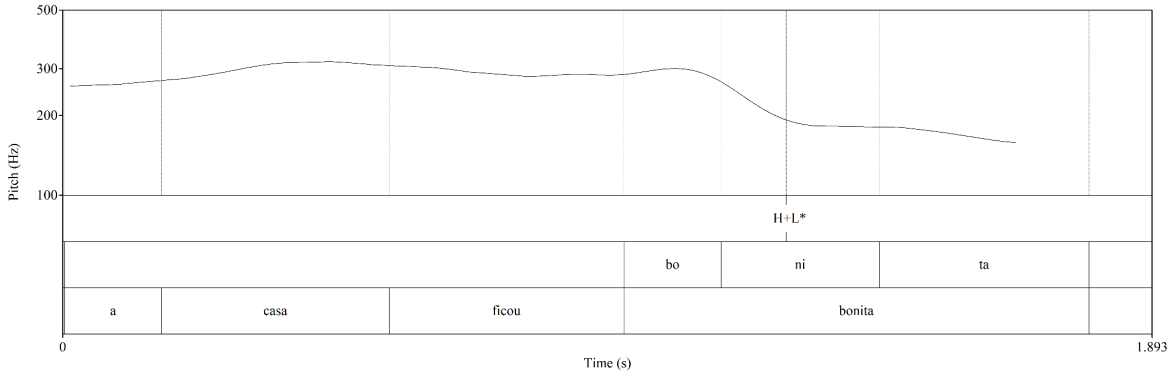
Tais evidências revelam a importância desse padrão fonológico para uma síntese de fala consistente e previsível. Nas Figuras de 57 a 63, são ilustrados exemplos em que o acento tonal descendente é atribuído às sílabas tônicas finais dos enunciados descritos.

Figura 57 – Acento tonal associado à última sílaba tônica do enunciado “Batata combina com peixe”, gerado pela voz pt-BR-Wavenet-C



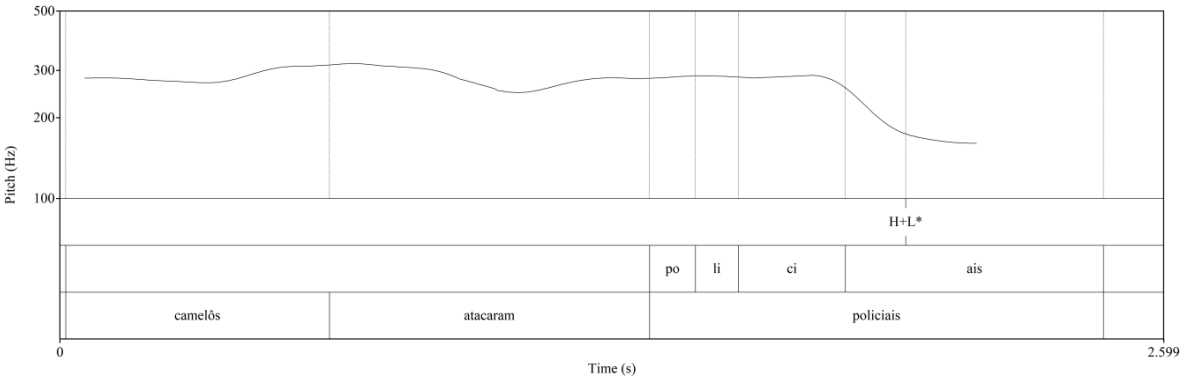
Fonte: Elaborado pelo autor (2025).

Figura 58 – Acento tonal associado à última sílaba tônica do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Standard-A



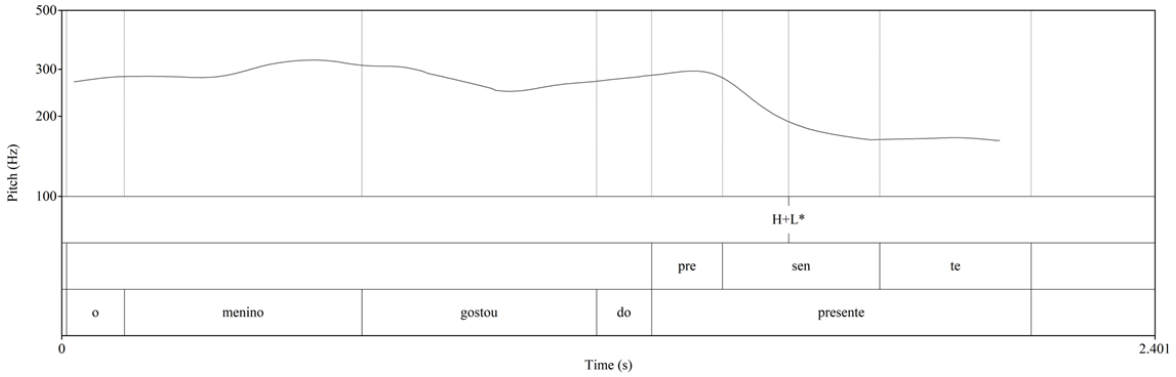
Fonte: Elaborado pelo autor (2025).

Figura 59 – Acento tonal associado à última sílaba tônica do enunciado “Camelôs atacaram policiais”, gerado pela voz pt-BR-Standard-A



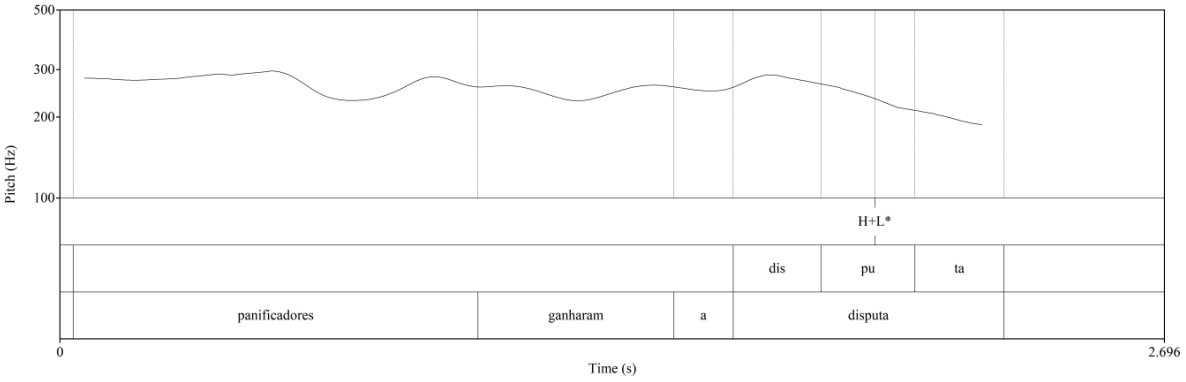
Fonte: Elaborado pelo autor (2025).

Figura 60 – Acento tonal associado à última sílaba tônica do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Standard-A



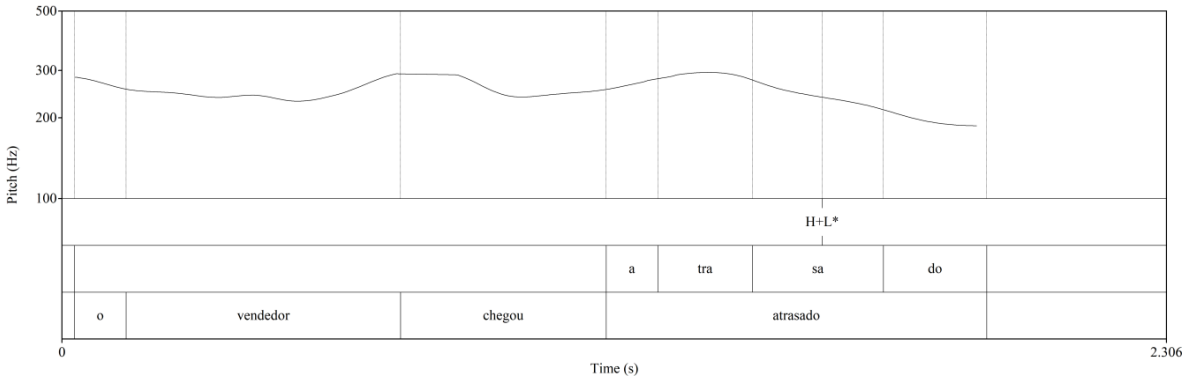
Fonte: Elaborado pelo autor (2025).

Figura 61 – Acento tonal associado à última sílaba tônica do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Wavenet-C



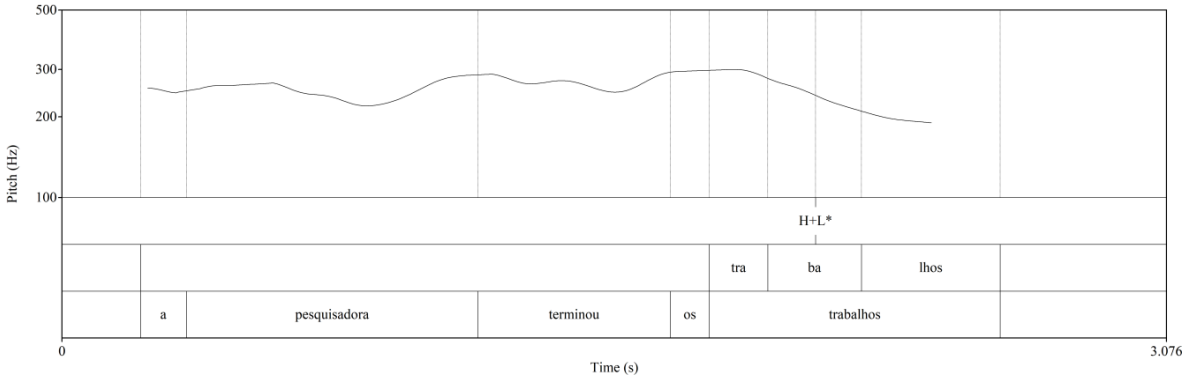
Fonte: Elaborado pelo autor (2025).

Figura 62 – Acento tonal associado à última sílaba tônica do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Wavenet-A



Fonte: Elaborado pelo autor (2025).

Figura 63 – Acento tonal associado à última sílaba tônica do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Neural2-C



Fonte: Elaborado pelo autor (2025).

Nos enunciados declarativos neutros produzidos pela tecnologia de síntese de fala do Google Cloud Text-to-Speech, a sílaba tônica saliente (Halliday, 1963, 1970; Cagliari, 2007), caracterizada, de um ponto de vista fonético, por uma variação da F_0 que “se sobrepõe a uma sílaba tônica em nível lexical” (Massini-Cagliari, 1992, p. 38), incide sobre a última sílaba tônica de cada sentença, o que corrobora a afirmação de o português brasileiro apresentar uma recursividade sintática à direita (Abaurre, 1996; Galves; Abaurre, 2002; Bisol, 2014). Em outras palavras, a sílaba tônica final dessas sentenças é o local do padrão melódico em que há uma significativa variação da F_0 (em particular, uma considerável descendência no contorno entoacional) em relação ao restante do enunciado (Massini-Cagliari, 1992). O contorno melódico descendente, manifestado pela queda da F_0, sinaliza a informação fonológica do término de uma frase entoacional no português brasileiro, que normalmente marca o fim de um enunciado e corresponde ao encerramento de uma unidade sintática e de sentido (Soncin; Tenani; Berti, 2017).

No que concerne à entoação declarativa neutra na gramática do português brasileiro, Tenani (2002) afirma que o acento principal recai sobre a última sílaba acentuada, à qual é associado o tom descendente H+L*. Segundo Massini-Cagliari e Cagliari (2012) e Massini-Cagliari (2017), o português brasileiro contemporâneo tem enunciados que carreiam, em termos de linha melódica, padrões previamente determinados pelo sistema linguístico, como é o caso das declarações, que sempre dispõem de um padrão entoacional descendente (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b).[98]

Kato (2005) observa que a tradição gerativista tem o objetivo de descrever a arquitetura mental responsável pela produção linguística. Esse ponto de vista propicia a interpretação dos resultados concernentes ao último acento tonal. Nos enunciados analisados, verifica-se a predominância do acento tonal descendente, característico de declarações neutras no português brasileiro. Essa consistência pode ser interpretada como uma demonstração de que a fala sintética do Google Cloud Text-to-Speech reproduz a principal característica prosódica da modalidade declarativa neutra no português brasileiro, que faz parte da gramática fonológica subjacente dos falantes.

Com base nas discussões acima, pode-se observar que, em relação ao acento tonal atribuído à última sílaba tônica dos enunciados declarativos neutros, há semelhanças entre a fala sintética e a natural. Nos dois tipos de fala, o acento principal da declaração neutra recai, de modo categórico, sobre a última sílaba tônica, à qual é atribuído o evento tonal H+L*, foneticamente expresso por uma descendência melódica no local mais saliente da curva entoacional.

A próxima subseção descreve e analisa o tom de fronteira atribuído à direita do contorno de entoação dos enunciados declarativos neutros.

4.2 TOM DE FRONTEIRA

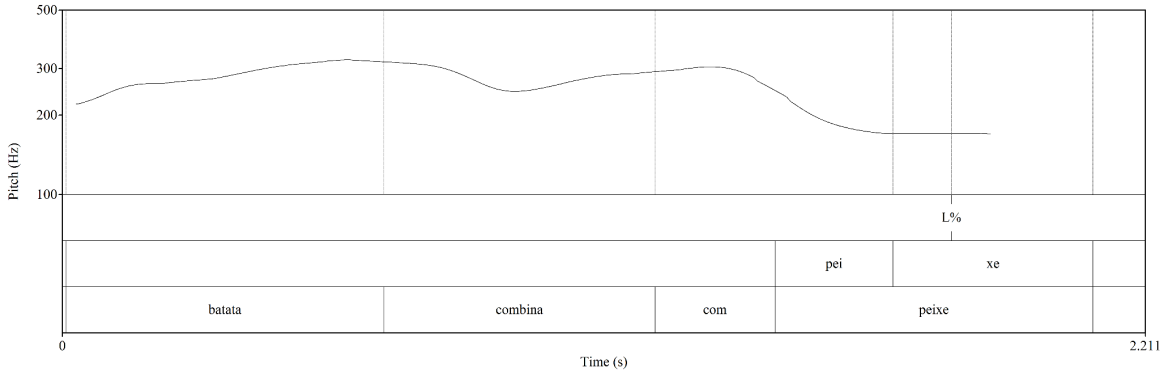
No conjunto de declarações neutras analisadas, o tom de fronteira atribuído à direita do contorno de entoação é L%. Há, no entanto, algumas ocorrências em que não ocorre a associação de tons à fronteira entoacional, como é discutido a seguir.

O evento tonal L%, atribuído à fronteira direita do contorno de entoação dos

[98] Para conhecer os aspectos da entoação do português arcaico, consultar Massini-Cagliari (2017, 2021, 2023a).

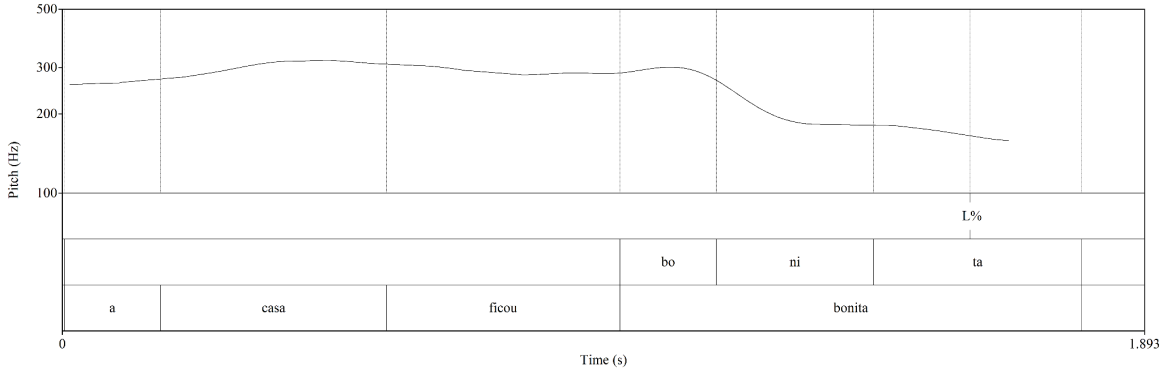
enunciados declarativos neutros, ocorre em 54 dados e caracteriza um *reset* (Serra, 2009). Esse tom de fronteira é foneticamente expresso por um maior abaixamento da curva melódica, em comparação com a sílaba tônica final. Nas Figuras de 64 a 69, são ilustradas ocorrências em que o tom de fronteira baixo é atribuído às sílabas postônicas finais dos enunciados, com exceção da sentença "Camelôs atacaram policiais".

Figura 64 – Tom de fronteira associado à direita do contorno de entoação do enunciado "Batata combina com peixe", gerado pela voz pt-BR-Standard-A



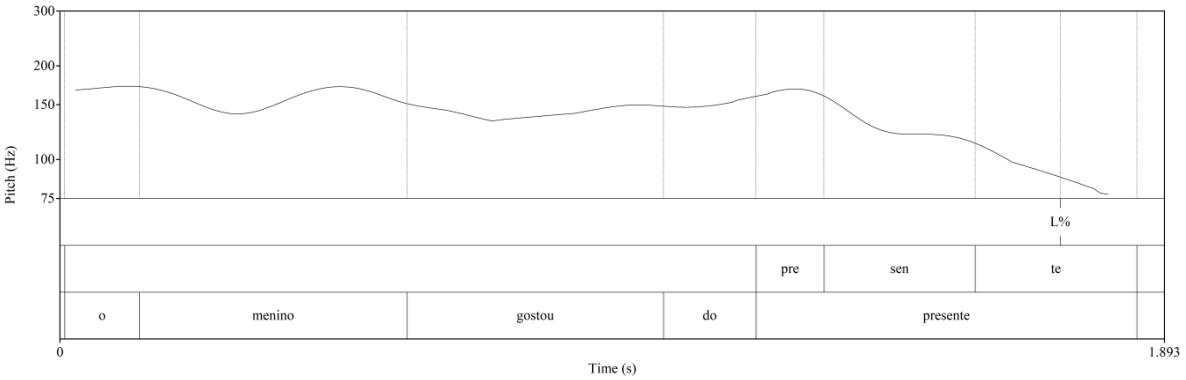
Fonte: Elaborado pelo autor (2025).

Figura 65 – Tom de fronteira associado à direita do contorno de entoação do enunciado "A casa ficou bonita", gerado pela voz pt-BR-Standard-A



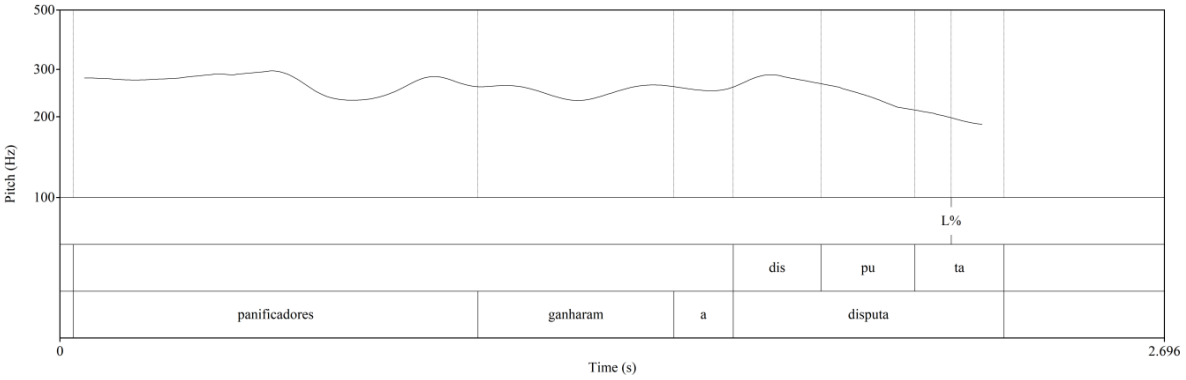
Fonte: Elaborado pelo autor (2025).

Figura 66 – Tom de fronteira associado à direita do contorno de entoação do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Standard-B



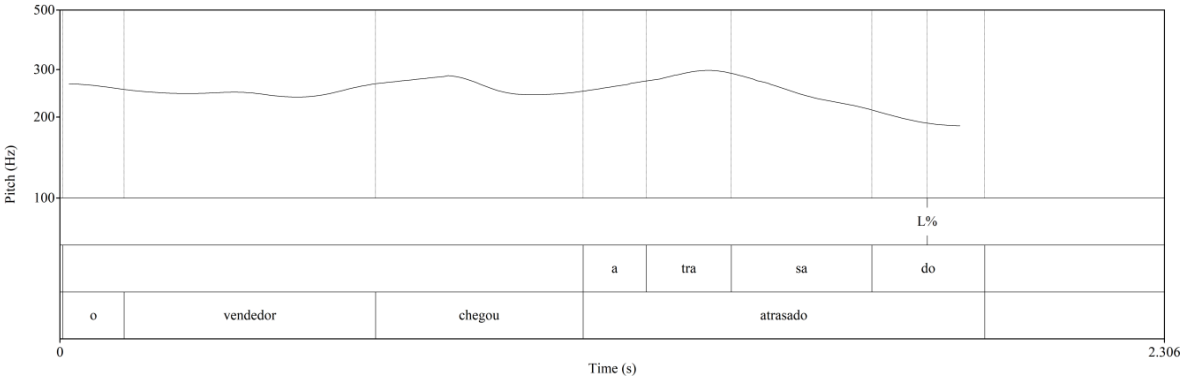
Fonte: Elaborado pelo autor (2025).

Figura 67 – Tom de fronteira associado à direita do contorno de entoação do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Wavenet-C



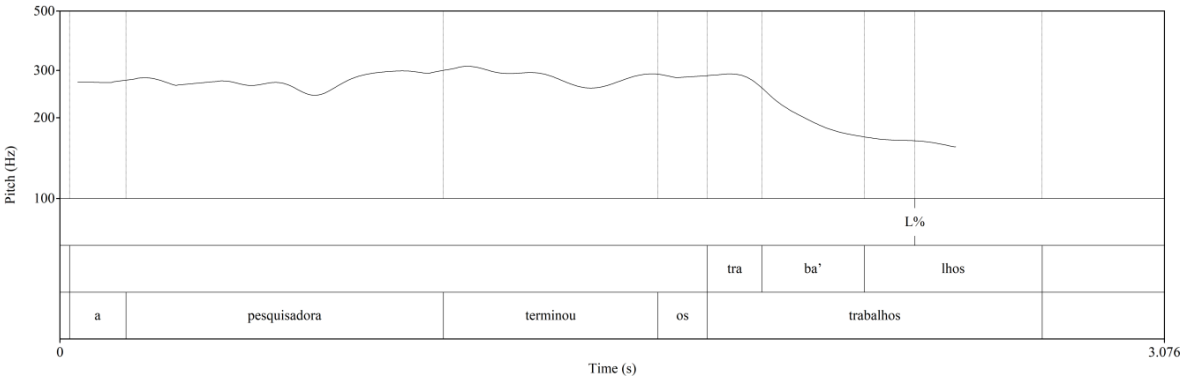
Fonte: Elaborado pelo autor (2025).

Figura 68 – Tom de fronteira associado à direita do contorno de entoação do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Standard-C



Fonte: Elaborado pelo autor (2025).

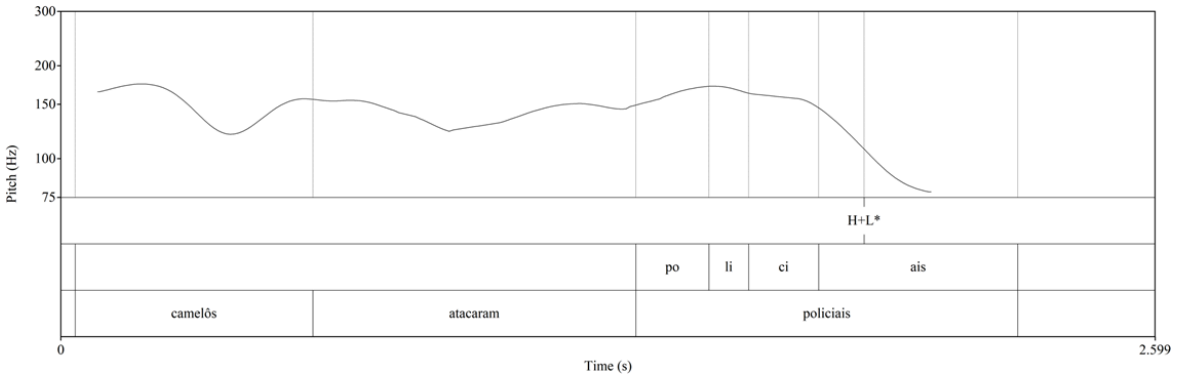
Figura 69 – Tom de fronteira associado à direita do contorno de entoação do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Neural2-A



Fonte: Elaborado pelo autor (2025).

Nas nove ocorrências do enunciado “Camelôs atacaram policiais”, não é verificada a ocorrência de um tom de fronteira baixo. Em virtude de a palavra “policiais” ser oxítona (com tonicidade na última sílaba), não há qualquer material fônico para a atribuição tonal após a sílaba tônica “ais” (Tenani, 2002). Na Figura 70, é exemplificada uma ocorrência do enunciado “Camelôs atacaram policiais”, em que há, na palavra fonológica “policiais”, apenas a associação do acento tonal H+L* à sílaba tônica “ais”.

Figura 70 – Acento tonal associado à última sílaba tônica do enunciado “Camelôs atacaram policiais”, gerado pela voz pt-BR-Standard-B



Fonte: Elaborado pelo autor (2025).

Embora não seja atribuído, no nível fonológico, um tom de fronteira à direita da frase entoacional referente ao enunciado “Camelôs atacaram policiais”, observa-se que, na manifestação fonética do acento tonal associado à sílaba tônica nuclear “ais”, há um maior declínio na variação da gama da F_0 (*pitch range variation*), de modo a indicar, em termos linguísticos, o término da sentença declarativa neutra. Em outras palavras, apesar de não ocorrer a atribuição do tom de fronteira baixo, um expediente acústico é implementado para

sinalizar o encerramento do enunciado.[99]

De acordo com os resultados, a ausência do tom de fronteira resulta de fatores linguísticos previsíveis. Quando determinadas condições não são atendidas, a aplicação do tom de fronteira deixa de ser realizada. A análise estatística confirma a regularidade da implementação do tom de fronteira baixo no sistema de síntese de fala, ou seja, sugere que esse resultado não é aleatório. Trata-se de uma análise indicativa da presença consistente do tom de fronteira baixo nos enunciados declarativos neutros do português brasileiro.[100]

Quanto à gramática do português brasileiro, Tenani (2002) atesta que, em sentenças declarativas neutras, ocorre um tom baixo à direita do contorno melódico, a não ser que a última sílaba tônica ocupe a posição final da frase entoacional. Tal fato decorre, segundo a autora, da ausência de material fônico para que esse tom seja implementado.

Conforme Kato (2002), a teoria gerativa prevê a existência de propriedades invariantes, embora certas variações possam decorrer da experiência linguística. Os resultados relativos ao tom de fronteira confirmam essa perspectiva, uma vez que evidenciam, por meio da análise fonética, a predominância de um maior abaixamento da curva melódica no término dos enunciados, em conformidade com os aspectos melódicos descritos nos dados de fala natural, mesmo nos casos em que não há a presença de material fônico para a implementação melódica. Essa correspondência é indicativa de que o sistema de síntese de fala examinado modela, a partir de dados humanos e de procedimentos estatísticos, uma característica fundamental da prosódia do português brasileiro, o que também reforça a eficácia dos princípios teóricos delineados pela teoria gerativa de Chomsky e Halle (1968) em relação à existência de uma gramática fonológica subjacente na fala natural.

Tendo descrito, respectivamente na subseção anterior e na subseção em pauta, o acento tonal atribuído à última sílaba tônica e o tom de fronteira associado à direita do contorno de entoação, é imprescindível que sejam feitas algumas considerações sobre o contorno entoacional nuclear do conjunto de enunciados declarativos neutros. Com exceção das nove ocorrências da sentença "Camelôs atacaram policiais", finalizada com uma palavra oxítona, o

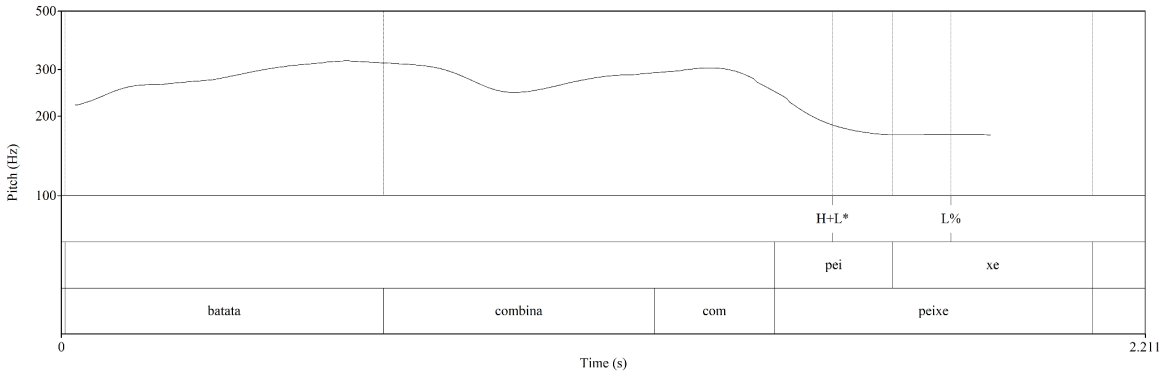
---

[99] Mesmo na ausência de material fônico para a implementação do tom de fronteira baixo, a declinação melódica ocorre por meio de movimentos locais da F_0 e de ajustes temporais finais. Esse mecanismo destaca a relação entre a Fonética, que fornece as pistas acústicas, e a Fonologia, que define as categorias melódicas. Esses domínios não são dicotômicos, mas complementares, e podem ser explorados pela síntese de fala.

[100] A variabilidade observada não corresponde à aleatoriedade, mas às possibilidades gramaticalmente permitidas pelo português brasileiro. Apesar da falta de material fônico suficiente para a produção do tom de fronteira baixo, o sistema utiliza uma variação fonética que preserva a descendência melódica e confirma a consistência da estrutura entoacional. Em palavras oxítonas finais, a língua admite a aplicação de recursos acústicos mínimos, como a diminuição local da F_0, para sinalizar a declinação típica da declaração neutra. Trata-se de uma variabilidade licenciada pelo português brasileiro, que não deve ser confundida com um ruído indesejado ou uma falha técnica do sistema de síntese de fala.

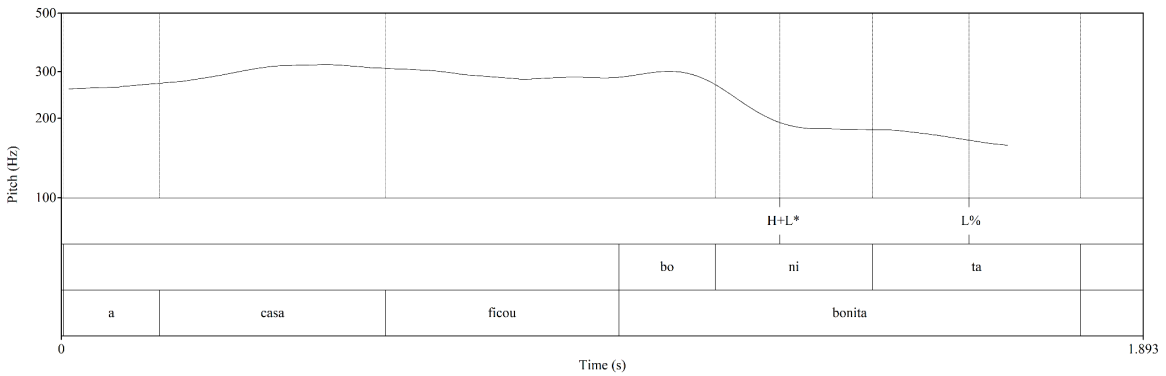
contorno entoacional nuclear do conjunto de dados da amostra é H+L* L%, que sinaliza, assim como na fala natural, a modalidade pragmática da declaração neutra (Cunha, 2000; Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). A ocorrência do contorno entoacional nuclear H+L* L% no Google Cloud Text-to-Speech não tem um caráter aleatório, tampouco é imprevisível ou irregular. O resultado evidencia que esse tipo de contorno é frequentemente utilizado na estrutura entoacional de enunciados declarativos neutros. A probabilidade de ocorrência aleatória desse contorno é mínima, pois ele se manifesta de maneira sistemática no sistema examinado. As Figuras de 71 a 76 ilustram exemplos em que o contorno entoacional nuclear dos enunciados é H+L* L%.

Figura 71 – Contorno entoacional nuclear do enunciado “Batata combina com peixe”, gerado pela voz pt-BR-Standard-A



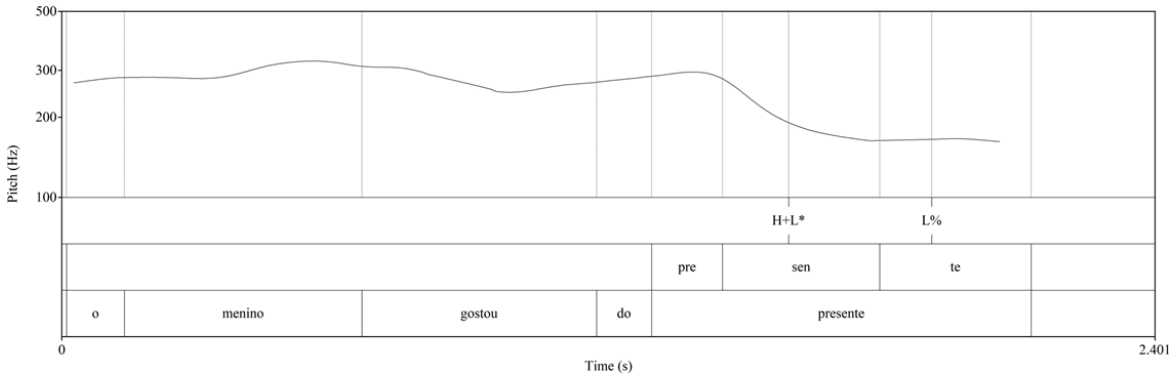
Fonte: Elaborado pelo autor (2025).

Figura 72 – Contorno entoacional nuclear do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Standard-A



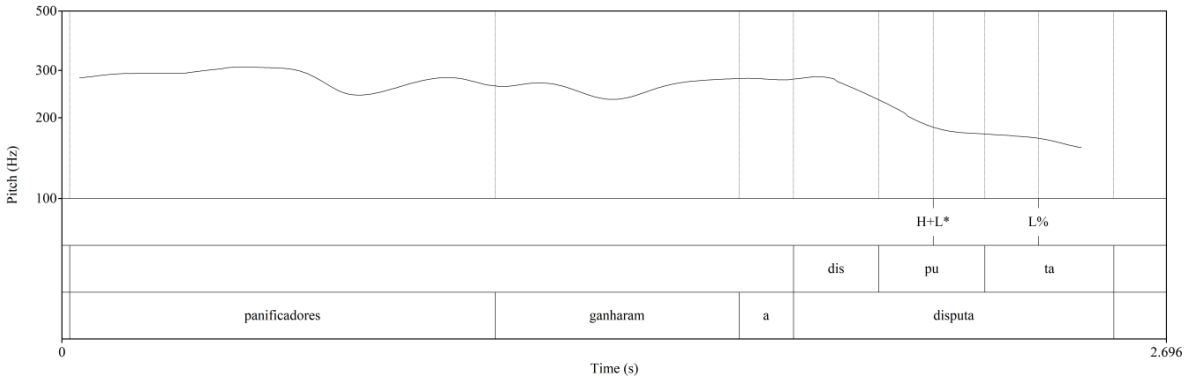
Fonte: Elaborado pelo autor (2025).

Figura 73 – Contorno entoacional nuclear do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Standard-A



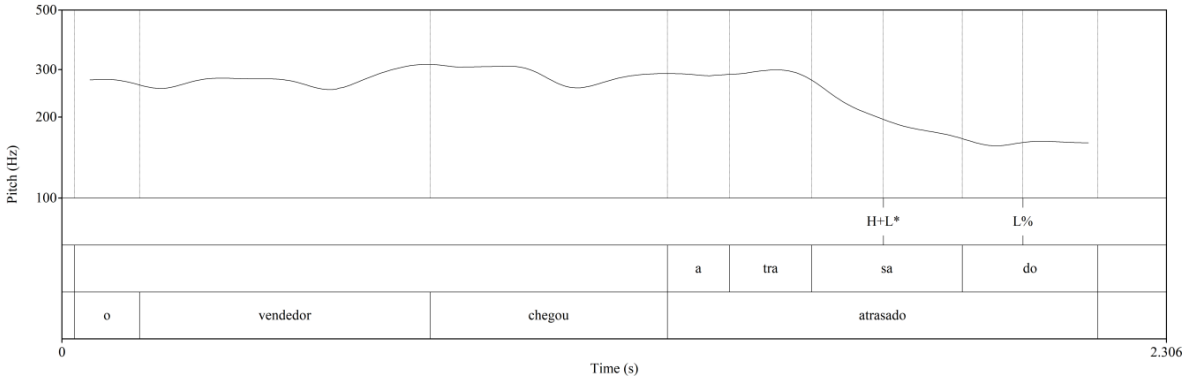
Fonte: Elaborado pelo autor (2025).

Figura 74 – Contorno entoacional nuclear do enunciado “Panificadores ganharam a disputa”, gerado pela voz pt-BR-Standard-A



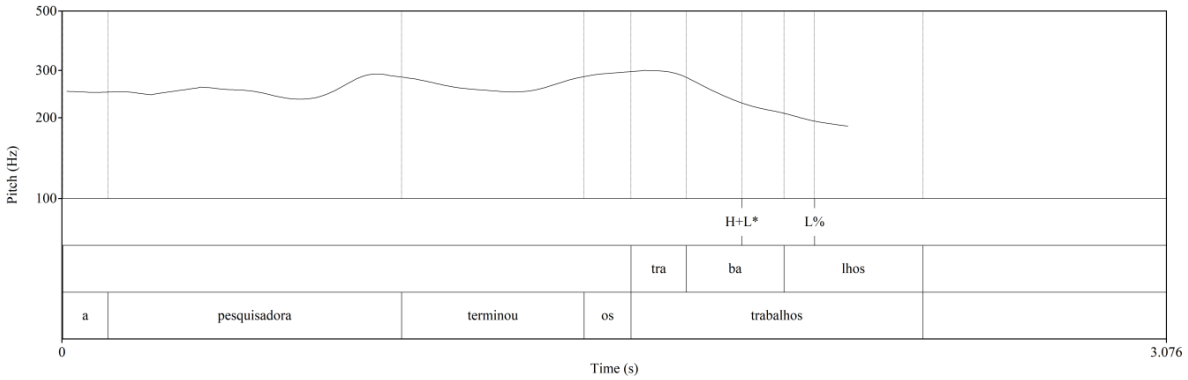
Fonte: Elaborado pelo autor (2025).

Figura 75 – Contorno entoacional nuclear do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-A



Fonte: Elaborado pelo autor (2025).

Figura 76 – Contorno entoacional nuclear do enunciado “A pesquisadora terminou os trabalhos”, gerado pela voz pt-BR-Standard-C



Fonte: Elaborado pelo autor (2025).

A partir das discussões anteriores, pode-se observar que, quanto ao tom de fronteira atribuído à direita do contorno de entoação, há semelhanças entre a fala sintética e a natural. Nos dois tipos de fala, quando há um material fônico para a implementação melódica, o tom de fronteira dos enunciados declarativos neutros é L%, cuja expressão, no nível fonético, dá-se por meio de um maior abaixamento da linha entoacional, em comparação com a melodia incidente na última sílaba tônica. Esse tom de fronteira, em união com o acento tonal analisado na seção anterior, forma o contorno entoacional nuclear H+L* L%, que mapeia o significado pragmático da declaração neutra na amostra examinada. Assim, pode-se observar que a tecnologia de síntese de fala do Google Cloud Text-to-Speech, além de eficiente na modelagem das propriedades estruturais e inerentes à entoação declarativa neutra do português brasileiro, é capaz de reproduzir certos elementos discursivos da enunciação.

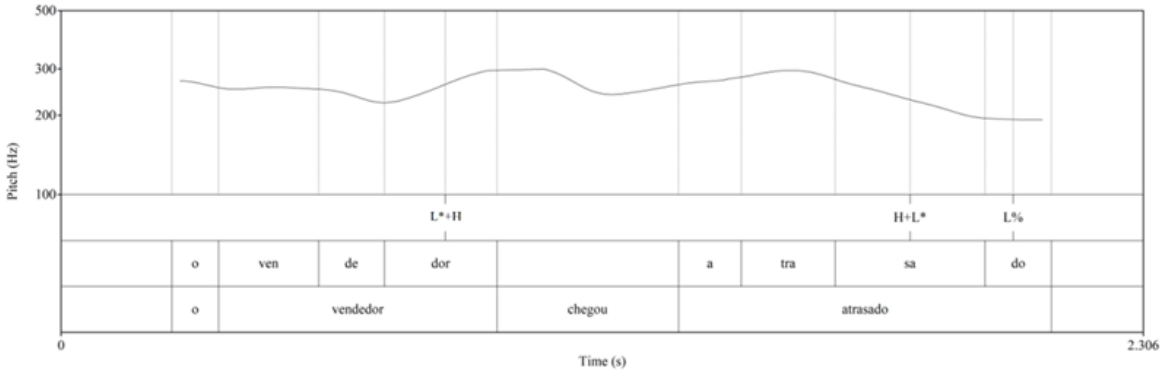
Na próxima subseção, são descritas as configurações entoacionais dos enunciados declarativos neutros. Não é apresentada uma amostra exaustiva de figuras e de informações, pois, nas subseções precedentes, são analisados, com detalhes e exemplificações, os principais eventos tonais associados à estrutura entoacional dos enunciados declarativos neutros.

4.3 CONFIGURAÇÕES ENTOACIONAIS

A configuração entoacional L*+H _____ H+L* L% é predominantes nos dados coletados, já que ocorre em 54 ocorrências analisadas, independentemente do tipo de enunciado escolhido. Refere-se ao padrão entoacional comum em frases afirmativas, marcado, no nível fonético, por um aumento gradual da F_0, que apresenta o valor máximo na última sílaba tônica do enunciado (Oliveira Jr., 2022), com a curva melódica descendente a partir

dessa sílaba. Na Figura 77, é exemplificado o contorno entoacional do enunciado “O vendedor chegou atrasado”, em que os acentos tonais ascendente e descendente são associados às sílabas tônicas “dor” e “sa”, e o tom de fronteira baixo é atribuído à postônica “do”.

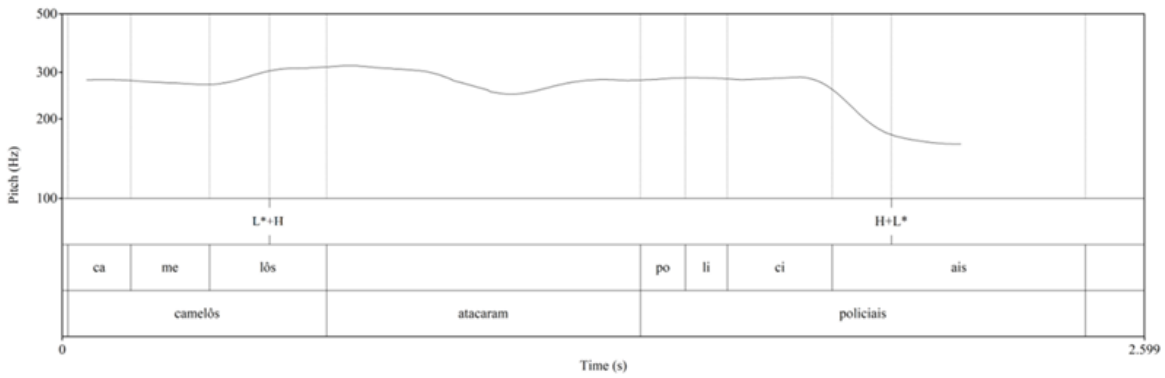
Figura 77 – Configuração entoacional do enunciado “O vendedor chegou atrasado”, gerado pela voz pt-BR-Neural2-C



Fonte: Elaborado pelo autor (2025).

A configuração entoacional L*+H _____ H+L* ocorre, por sua vez, nas nove realizações do enunciado “Camelôs atacaram policiais”. Na Figura 78, é ilustrado o contorno melódico do referido enunciado, em que os acentos tonais ascendente e descendente são atribuídos às sílabas tônicas “lôs” e “ais”. Como não há a presença de um material fônico após a sílaba tônica “ais”, visto que “policiais” consiste em uma palavra oxítona, é ausente a efetivação do tom de fronteira baixo (Tenani, 2002). Ainda assim, aplica-se um recurso acústico de diminuição da F_0 para sinalizar a conclusão do enunciado.

Figura 78 – Configuração entoacional do enunciado “Camelôs atacaram policiais”, gerado pela voz pt-BR-Standard-A



Fonte: Elaborado pelo autor (2025).

Os achados que apontam para a ocorrência desses padrões entoacionais nas saídas do Google Cloud Text-to-Speech devem ser interpretados como uma evidência de que o sistema reproduz as regularidades melódicas coerentes com a prosódia do português brasileiro na modalidade declarativa neutra. Essa correspondência não é uma reprodução acústica superficial, mas representa uma generalização baseada na distribuição de regras fonológicas entre os elementos da estrutura prosódica, derivada dos dados linguísticos humanos utilizados no treinamento do sistema computacional.[101] Assim, as configurações melódicas descritas também são encontradas na fala natural, conforme o profícuo diálogo estabelecido, nas subseções precedentes, com o estudo de Tenani (2002) e, em momentos oportunos, com as demais pesquisas referentes ao português brasileiro. Essa e outras características entoacionais dos enunciados declarativos neutros em contextos de fala sintética e natural são resumidas no Quadro 3.

Quadro 3 – Principais características entoacionais dos enunciados declarativos neutros em contextos de fala sintética e natural

Aspectos entoacionais básicos dos enunciados declarativos neutros	**Fala sintética**	**Fala humana**
Alta densidade tonal	+	+
Acento tonal na primeira sílaba tônica		
L*+H	+	+
Acento tonal na última sílaba tônica		
H+L*	+	+
Tom de fronteira (quando há material fônico)		
L%	+	+
Contorno entoacional nuclear		
H+L* L%	+	+
Configurações entoacionais		
L*+H H+L* L%	+	+
L*+H H+L*	+	+

Fonte: Elaborado pelo autor (2025) com o auxílio do ChatGPT/OpenAI (GPT-4o mini).[102]

Com base na análise linguística da síntese de fala no português brasileiro, constata-se que o Google Cloud Text-to-Speech consegue reproduzir, com precisão, os padrões entoacionais típicos da fala natural para os enunciados declarativos neutros. As características melódicas observadas, tais como a presença de acento tonal no começo e no término dos

[101] Ao defender a existência de um módulo especializado para a gramática, Kato (1999) contribui para a interpretação dos resultados referentes à configuração entoacional. O padrão melódico predominante, caracterizado por uma ascendência inicial seguida de uma descendência final, é uma evidência de que a entoação sintética gera as particularidades prosódicas do português brasileiro para as declarações neutras.
[102] Disponível em: https://openai.com/. Acesso em: 9 de agosto de 2024.

enunciados, o tom de fronteira, o contorno entoacional nuclear e a configuração melódica, são consistentes com a fala natural. Esse resultado evidencia que a distribuição dos tons gramaticais segue regularidades estatísticas no sistema de síntese de fala, com uma alta probabilidade de ocorrência na classe dos enunciados declarativos neutros do português brasileiro.[103] A hipótese de aleatoriedade não se mostra satisfatória para explicar esses aspectos entoacionais, uma vez que eles aparecem com mais frequência do que o esperado em um arranjo aleatório. Como consequência, o Google Cloud Text-to-Speech mantém a consistência e a regularidade na estruturação dos padrões entoacionais dos enunciados declarativos neutros, o que confirma, para os dados descritos, a eficácia do sistema na modelagem computacional da prosódia do português brasileiro e na capacidade de reproduzir as propriedades melódicas da gramática fonológica da fala natural, existente, de maneira inata e abstrata, na cognição humana, segundo a perspectiva gerativista (Chomsky; Halle, 1968).

A constatação de que os contornos entoacionais declarativos neutros da fala sintética gerada pelo sistema do Google Cloud Text-to-Speech se alinham, no nível fonológico, àqueles observados na fala natural não implica, contudo, uma equivalência fonético-acústica entre os dois tipos de produção, tampouco uma correspondência exata no plano perceptivo.[104] Embora a modelagem computacional reproduza, com regularidade, os padrões de distribuição tonal característicos de enunciados declarativos neutros do português brasileiro, não se pode afirmar que as medidas da F_0, assim como sua interação com outros parâmetros acústicos, manifestem-se de maneira idêntica à observada na fala humana. A confirmação desse aspecto demanda uma investigação empírica, a partir de análises fonético-acústicas detalhadas e de testes perceptivos, ambos respaldados por procedimentos estatísticos mais complexos.[105]

Os resultados obtidos demonstram que a fala sintética, embora baseada em métodos estatísticos, reproduz, para as declarações neutras, os padrões entoacionais consistentes e comparáveis aos da fala natural do português brasileiro (Cunha, 2000; Frota; Vigário, 2000;

---

[103] As três vozes analisadas têm a mesma organização gramatical na entoação declarativa neutra. Os padrões melódicos permanecem estáveis e confirmam que as diferenças entre os modelos Standard, Neural2 e WaveNet se concentram em características técnicas, sem afetar a estrutura linguística.

[104] Por razões técnicas, há diferenças audíveis entre as vozes, mas elas se restringem ao nível fonético e implementacional. Essas diferenças não alteram a estrutura gramatical da entoação declarativa neutra, que permanece consistente em todos os casos.

[105] A distinção entre a fala humana e a sintética pode surgir tanto de características dos segmentos quanto dos elementos prosódicos. Para que se compreenda a contribuição de cada fator, deve-se propor a realização de experimentos psicoacústicos que considerem diferentes condições, como manipular a prosódia com os segmentos preservados, manipular os segmentos com a prosódia preservada e aplicar tarefas de avaliação de naturalidade ou de identificação. É importante investigar se os sinais acústicos com uma prosódia gramaticalmente coerente seriam percebidos como naturais por ouvintes humanos, mesmo com a presença de pequenas variações nos segmentos. As diferenças percebidas nos segmentos ou na prosódia podem sinalizar a origem sintética do sinal, o que reforça a necessidade de estudos futuros para identificar os parâmetros determinantes para a percepção de naturalidade.

Tenani, 2002; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). Essa evidência não significa que o sistema de síntese de fala dispõe de uma gramática fonológica inata, mas mostra que os padrões gramaticais representados na cognição linguística humana se mostram suficientemente robustos para serem observados também em modelos estatísticos de larga escala durante a produção de enunciados declarativos neutros. Nesse sentido, a análise desenvolvida reforça a pertinência de uma abordagem gerativista, ao evidenciar que os contornos entoacionais do português brasileiro reproduzem a sistematicidade melódica e as restrições fonológicas inerentes à competência linguística dos falantes em declarações neutras.[106]

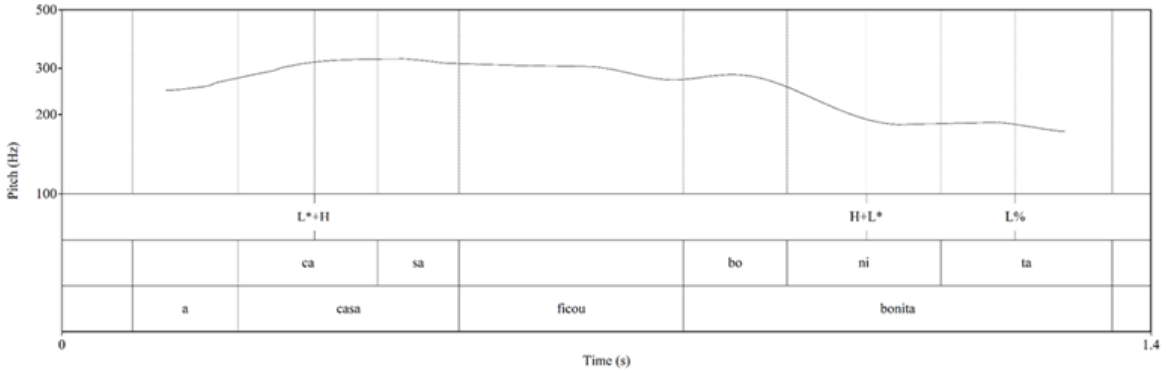
Testa-se, em caráter exploratório, um modelo de voz adicional, conhecido como Chirp 3, recentemente disponibilizado pelo Google Cloud Text-to-Speech, que não compõe o conjunto principal de análises. As amostras declarativas neutras têm a mesma estrutura entoacional já documentada nas demais vozes e se diferenciam somente por aspectos técnicos.[107] Nas Figuras 79 e 80, apresentam-se duas representações do contorno entoacional produzidas por esse modelo, correspondentes a uma voz feminina e a uma voz masculina. Apesar de todos os enunciados serem produzidos de forma robusta, são exibidos apenas dois exemplos, cada um com um tipo de voz, para fornecer uma visão geral da organização melódica nesse modelo e evitar o excesso de dados de um recurso que não faz parte do escopo principal da pesquisa. Esses achados preliminares ratificam a ideia de que a organização fonológica da entoação declarativa neutra permanece coerente no sistema de síntese de fala estudado e destacam a necessidade de uma atenção específica a novos modelos em trabalhos

[106] A semelhança entre a fala natural e a fala sintética decorre do funcionamento dos modelos de síntese de fala atuais (van den Oord *et al.*, 2016; Casanova, 2019, 2022; Casanova; Shulby; Aluísio, 2021; Casanova *et al.*, 2021, 2022, 2024; Barbosa, 2022). Eles costumam incluir uma etapa linguística, responsável pelo pré-processamento, pela conversão de grafemas em fonemas e pela análise morfossintática, além de um módulo prosódico, que codifica as regularidades rítmicas e entoacionais, e de uma etapa de geração neural, que modela, com eficiência, as trajetórias e a forma de onda. Essa arquitetura computacional faz com que o sistema seja capaz de processar as categorias prosódicas de maneira sistemática e de reproduzir os padrões melódicos comuns, como a entoação declarativa neutra do português brasileiro.

[107] Uma das alternativas para o pesquisador é ilustrar ocorrências de outras modalidades de enunciados, como os interrogativos e os imperativos. Contudo, para não se restringir a uma nova classe de enunciados, o que implica a exclusão das demais, e para evitar a ampliação do texto com discussões que se afastam do escopo principal da pesquisa, decide-se por ilustrar enunciados declarativos neutros de um modelo de voz recentemente lançado. Os enunciados declarativos neutros constituem o enfoque do estudo e as seções anteriores oferecem uma fundamentação teórica detalhada sobre essa classe no português brasileiro. Assim, o leitor pode acompanhar a análise de modo esclarecedor e comparar os dados apresentados com uma discussão teórica apropriada. As descrições entoacionais referentes a outras classes de enunciados permanecem como um possível tema para estudos subsequentes, específicos para essas modalidades, com um embasamento teórico-descritivo particular.

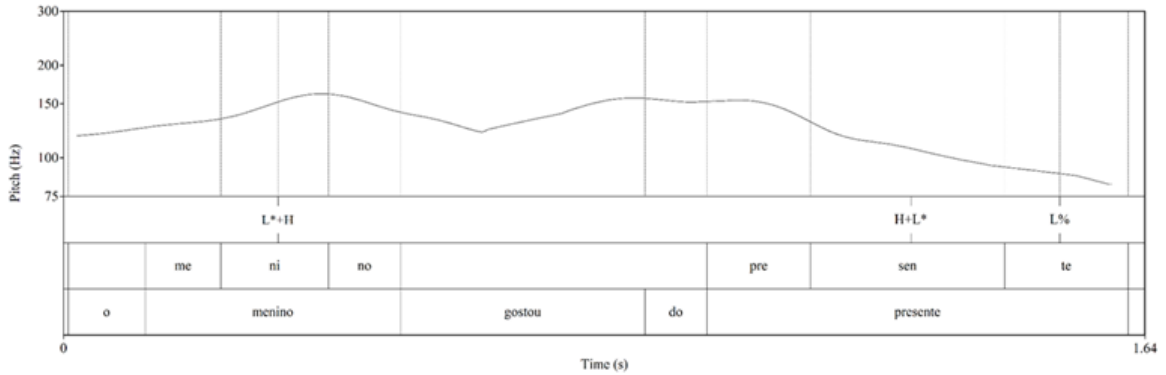
posteriores.

Figura 79 – Contorno entoacional do enunciado “A casa ficou bonita”, gerado pela voz pt-BR-Chirp3-HD-Achernar



Fonte: Elaborado pelo autor (2025).

Figura 80 – Contorno entoacional do enunciado “O menino gostou do presente”, gerado pela voz pt-BR-Chirp3-HD-Puck



Fonte: Elaborado pelo autor (2025).

Como analisado adicionalmente, a similaridade entre a fala sintética e a natural em declarações neutras evidencia a possibilidade de as representações entoacionais serem concebidas como princípios computáveis, ou seja, estruturas que podem ser modeladas por programas treinados com *corpora* linguísticos. Esse resultado tem implicações epistemológicas, pois reforça a ideia de que as categorias fonológicas representam as regularidades distribuídas na linguagem, e não meros reflexos de mecanismos físicos particulares. Em termos práticos, permite-se a verificação de hipóteses prosódicas em novos meios (dados de fala sintética) que expandem o escopo empírico da Fonética e da Fonologia.

O avanço da tecnologia de síntese de fala requer uma atualização regular. Durante a elaboração da pesquisa, são lançados outros modelos de voz com uma entoação declarativa neutra igualmente apropriada. Essa evolução confirma que a tecnologia de síntese de fala não

deixa de considerar tanto os elementos fonológicos quanto os aspectos pragmático-discursivos do português brasileiro para fazer com que os áudios sejam mais naturais. O modelo adicional de voz mencionado anteriormente, o Chirp 3, comprova a reprodutibilidade da entoação declarativa neutra no Google Cloud Text-to-Speech. Ademais, os novos indícios reforçam a validade da interpretação fonológica feita na pesquisa e estimulam a realização de futuras investigações sobre as propriedades melódicas das variadas modalidades comunicativas.

A seção subsequente é dedicada à conclusão, em que são apresentados os principais resultados do estudo e as contribuições do trabalho para o campo da prosódia da fala sintética.

5 CONCLUSÃO

A pesquisa se propõe a caracterizar as propriedades entoacionais de enunciados declarativos neutros gerados por um sistema de conversão de texto escrito em fala audível baseado em IA, conhecido como Google Cloud Text-to-Speech. O estudo também aborda a comparação entre a entoação sintética e a natural do português brasileiro. Para tanto, os eventos tonais atribuídos ao contorno entoacional desses enunciados são analisados, com o objetivo de determinar o padrão de entoação predominante e verificar a compatibilidade da fala sintética com a prosódia da variedade brasileira da língua portuguesa. O trabalho, fundamentado na Fonologia Entoacional Autossegmental e Métrica (Ladd, 1996, 2008) e na Fonologia Prosódica (Nespor; Vogel, 1986, 2007), é feito com o auxílio da inspeção auditiva do pesquisador e da versão 6.4.04 do *software* Praat (Boersma; Weenink, 2024) e segue as diretrizes do sistema P-ToBI (Frota; Oliveira, P.; Cruz; Vigário, 2015). Em seguida, os resultados do estudo são comparados com os descritos por Tenani (2002) para a entoação declarativa neutra em contextos de fala natural, bem como com outros trabalhos acerca da entoação do português brasileiro (Cunha, 2000; Frota; Vigário, 2000; Cagliari, 2007; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Serra, 2009; Massini-Cagliari; Cagliari, 2012; Silvestre, 2012; Córdula, 2013; Frota *et al.*, 2015; Frota; Moraes, 2016; Castelo, 2016; Massini-Cagliari, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b).[108]

Os resultados da pesquisa indicam que a fala sintética analisada dispõe, no escopo da declaração neutra, de uma estrutura entoacional consistente com a gramática prosódica do português brasileiro. A investigação também identifica um padrão de entoação prototípico, caracterizado por uma ascendência tonal no começo e uma descendência melódica no término dos enunciados, e reforça a tendência de alta densidade tonal nessa variedade linguística, atestada por Frota *et al.* (2015), Frota e Moraes (2016) e Fernandes-Svartman (2024a, 2024b) em dados de fala natural. Ademais, a atribuição de acentos tonais ocorre de acordo com a organização prosódica do português brasileiro e destaca a importância da palavra fonológica nesse fenômeno linguístico (Tojeira-Ramos; Massini-Cagliari, no prelo).

[108] Deve-se salientar que os resultados e as generalizações discutidos no trabalho se referem somente à entoação declarativa neutra do português brasileiro, observada em períodos sintáticos simples gerados por um sistema de síntese de fala. A investigação não contempla a variação melódica de diversas modalidades enunciativas, tampouco de estruturas sintáticas mais complexas, de diferentes línguas e variedades dialetais, ou de outros sistemas de processamento computacional de fala. Esses domínios sugerem oportunidades para pesquisas futuras, que podem ampliar o alcance analítico do estudo e propiciar uma avaliação mais abrangente das propriedades entoacionais da fala sintética em distintos contextos linguísticos e tecnológicos.

Outro aspecto relevante que merece atenção é a semelhança entre os eventos tonais observados na fala sintética e aqueles identificados por Tenani (2002) na fala natural. Essa correspondência sugere que o Google Cloud Text-to-Speech pode reproduzir, com precisão, a estrutura entoacional do português brasileiro no tocante à declaração neutra. A capacidade de reproduzir os padrões entoacionais detectados na fala natural para essa modalidade enunciativa, além de demonstrar um avanço significativo nas tecnologias de síntese de fala baseadas em IA, valida o potencial desse tipo de sistema para aplicações práticas, tais como laringe eletrônica (ou eletrolaringe), assistência virtual, educação midiática, leitura de tela e acessibilidade para pessoas com deficiência visual.

De acordo com as informações acima, confirmam-se as hipóteses do trabalho, que são retomadas e discutidas conjuntamente. Em primeiro lugar, os enunciados declarativos neutros sintetizados pelas vozes Standard, Neural2 e WaveNet compartilham o mesmo esqueleto entoacional da fala natural no português brasileiro, com o acento tonal pré-nuclear ascendente e o nuclear descendente, seguido do tom de fronteira baixo. Em segundo lugar, a possível diferença de pronúncia entre os modelos de voz é de caráter técnico, sem que haja uma alteração no padrão de atribuição dos eventos tonais. Além disso, mesmo quando a fronteira direita é ausente para a implementação do tom baixo, como nas palavras oxítonas, que contam apenas com o acento nuclear descendente, o encerramento dos enunciados recebe uma sinalização fonética, pois são detectadas pistas acústicas suficientes para recuperar a diminuição gradual da F_0 e manter a consistência do contexto discursivo-pragmático, de modo semelhante ao que acontece na fala natural. Ratifica-se, assim, a necessidade de uma abordagem que integre a Fonética e a Fonologia para o estudo do módulo prosódico de sistemas de síntese de fala.[109]

O estudo contribui para o avanço do conhecimento em diferentes áreas. No campo da Fonética e da Fonologia, o trabalho amplia a compreensão da entoação declarativa neutra do português brasileiro ao analisar como ela se realiza em um contexto sintético e possibilita uma reflexão sobre a interface entre a prosódia e a tecnologia de fala. No domínio da Linguística Computacional, a pesquisa demonstra que os modelos de síntese de fala do Google Cloud Text-to-Speech já apresentam a capacidade de reproduzir, ao menos no caso dos enunciados declarativos neutros, os padrões entoacionais da fala natural, o que indica um avanço na

[109] A primeira hipótese não seria confirmada se os eventos tonais identificados não tivessem sido descritos na literatura prosódica sobre a entoação do português brasileiro. A segunda hipótese, por outro lado, não seria confirmada se outro padrão entoacional, diferente do atestado, fosse predominante no conjunto de enunciados declarativos neutros sintéticos. Por sua vez, a terceira hipótese não seria confirmada se não tivessem sido encontradas semelhanças entoacionais entre a fala sintética e a natural no tocante às declarações neutras da variedade brasileira do português.

modelagem prosódica. Ademais, a análise dos dados reforça a importância da integração de abordagens fonológicas e acústicas para a compreensão da prosódia da fala sintética.

Sob uma perspectiva tecnológica, em que se encontram, por exemplo, a Ciência da Computação, a Engenharia Eletrônica e a Engenharia Elétrica, as evidências encontradas no trabalho são pertinentes para que os sistemas de processamento de fala sejam aperfeiçoados. Além disso, os resultados podem subsidiar a otimização de modelos computacionais de prosódia, para que a síntese de fala em outras tecnologias digitais se aproxime cada vez mais da produção humana.

As descobertas da pesquisa fornecem um ponto de partida para uma série de investigações que têm o potencial de aprofundar o estudo da entoação sintética no português brasileiro e das aplicações que ela apresenta em diferentes contextos sociais. Um possível projeto futuro é a análise da entoação sintética em diferentes tipos de enunciados, tais como os interrogativos, os imperativos e os que dispõem de variação pragmática decorrente de focalização prosódica. Essas modalidades de enunciados exibem características entoacionais distintas (Gonçalves, 1998; Cagliari, 2007; Fernandes, 2007b; Tenani; Fernandes-Svartman, 2008; Moraes, 2008; Truckenbrodt; Sandalo; Abaurre, 2009; Córdula, 2013; Massini-Cagliari; Cagliari, 2012; Frota *et al.*, 2015; Frota; Moraes, 2016) e podem fornecer informações valiosas sobre a capacidade da fala sintética de reproduzir diversas nuances prosódico-discursivas.

Outra vertente promissora envolve a percepção da entoação sintética por falantes proficientes, a fim de compreender de que modo fatores como a inteligibilidade e a aceitabilidade prosódica são avaliados pelos ouvintes. Com o avanço dos modelos de IA, os estudos futuros ainda têm a possibilidade de explorar as estratégias de aprimoramento da prosódia sintética, por meio da incorporação de recursos linguístico-computacionais mais sofisticadas, com o intuito de aumentar a naturalidade da fala gerada por IA.

Os resultados da pesquisa fornecem subsídios empíricos à defesa de que a entoação declarativa neutra do português brasileiro, no Google Cloud Text-to-Speech, não se caracteriza como um mero fenômeno acústico ou computacional, mas evidencia os padrões estruturais que organizam a prosódia dessa variedade linguística. A correspondência entre os contornos entoacionais da fala sintética e da natural demonstra que a modelagem prosódica no sistema de síntese de fala estudado reflete os padrões fonológicos da fala humana para a declaração neutra. A regularidade na atribuição de acentos tonais, a alta densidade tonal e a configuração dos contornos melódicos na modalidade enunciativa descrita ratificam que a entoação pode resultar de um sistema de regras fonológicas organizado na cognição dos seres

humanos, cujos registros servem de fundamento para o sistema computacional. Os modelos de síntese de fala do Google Cloud Text-to-Speech, ao utilizarem os dados extraídos de gravações humanas, produzem os resultados melódicos coerentes com os padrões entoacionais característicos da declaração neutra no português brasileiro. Por conseguinte, a capacidade da tecnologia em gerar os contornos melódicos naturais para esses enunciados demonstra que as saídas acústicas dos modelos computacionais manifestam os padrões derivados de uma gramática fonológica humana, discutida por Chomsky e Halle (1968), o que fortalece a proposição de que a entoação é capaz de ser determinada por um conjunto de princípios sistemáticos (Pierrehumbert, 1980; Beckman; Pierrehumbert, 1986; Pierrehumbert; Beckman, 1988; Ladd, 1996, 2008).[110]

Os trabalhos de Kato (1995, 1997, 1999, 2001, 2002, 2005), alinhados à tradição gerativista encontrada em estudos como os de Chomsky (1957, 1965, 1986, 1993), colaboram para o entendimento da língua como um conhecimento interno e modular, estruturado por princípios comuns às línguas naturais. Essa concepção ajuda a explicar os resultados da pesquisa, pois a prosódia sintética analisada apresenta os padrões entoacionais compatíveis com aqueles observados na fala natural para a declaração neutra no português brasileiro. Sob essa ótica, os avanços da síntese de fala baseada em IA sugerem que a representação da entoação declarativa neutra preserva, no Google Cloud Text-to-Speech, a distribuição melódica prevista pela gramática fonológica do português brasileiro.[111]

A relevância da palavra fonológica (e não da morfológica) na organização dos acentos tonais, observada nos arquivos sonoros de fala sintética (Tojeira-Ramos; Massini-Cagliari, no prelo), corrobora a ideia de que domínios prosódicos específicos orientam a atribuição de proeminência melódica na declaração neutra do português brasileiro. Trata-se de uma discussão que encontra respaldo na literatura acadêmica sobre fala natural (Frota; Vigário, 2000; Tenani, 2002; Fernandes, 2007a, 2007b; Tenani; Fernandes-Svartman, 2008; Vigário; Fernandes-Svartman, 2010; Toneli, 2014; Toneli; Vigário; Abaurre, 2014;

---

[110] Sugere-se que a manutenção da estrutura gramatical da entoação declarativa neutra em ambientes computacionais, em relação à competência fonológica de um falante proficiente do português brasileiro, beneficia o processamento cognitivo e a compreensão dos ouvintes, em conjunto com outros fatores linguísticos e discursivos. Esse apontamento, porém, deve ser confirmado em futuros experimentos fonético-fonológicos, psicolinguísticos e neurolinguísticos.

[111] Embora se observe, do ponto de vista gramatical vinculado ao Gerativismo, uma semelhança fonológica categórica entre a entoação da fala sintética gerada pelo Google Cloud Text-to-Speech e a da fala natural do português brasileiro em enunciados declarativos neutros de períodos sintéticos simples, admite-se que um estudo estritamente fonético e estatístico, com uma validação experimental por um conjunto de avaliadores humanos, possa identificar alguma diferença acústica ou perceptual em relação à fala natural, alinhada ou não a outros parâmetros prosódicos e segmentais. Trata-se de uma sugestão de desdobramento para investigações subsequentes.

Fernandes-Svartman; Romano, 2017; Toneli; Abaurre; Vigário, 2018; Fernandes-Svartman, 2024a, 2024b). Além disso, esse entendimento indica que a estrutura prosódica constitui uma parte do sistema de sons da língua, interage com outros níveis gramaticais e obedece a regras específicas (Selkirk, 1984; Nespor; Vogel, 1986, 2007). Dessa maneira, o estudo busca enriquecer o debate teórico sobre a existência de uma gramática para a prosódia e propõe que os recentes avanços no campo da IA devem servir não somente como mecanismos para a análise do componente prosódico, mas também como fonte de novas evidências empíricas de que a estrutura entoacional adere a princípios simbólicos e organizados.

A pesquisa contribui para os Objetivos de Desenvolvimento Sustentável (ODS), estabelecidos pela Organização das Nações Unidas (ONU) para a Agenda 2030, ao evidenciar como o estudo da entoação declarativa neutra do português brasileiro, em um sistema de síntese de fala, pode resultar em benefícios sociais, pedagógicos e técnicos. No âmbito do ODS 3 (“Saúde e Bem-Estar”), os resultados fornecem subsídios para o desenvolvimento de tecnologias assistivas mais naturais e eficazes, capazes de apoiar a comunicação em contextos hospitalares ou de reabilitação. Em relação ao ODS 4 (“Educação de Qualidade”), o trabalho favorece a criação de leitores automáticos e de materiais pedagógicos cuja entoação adequada potencializa a aprendizagem e amplia o acesso às práticas educacionais inclusivas. No ODS 9 (“Indústria, Inovação e Infraestrutura”), a pesquisa ressalta a relevância da integração entre o conhecimento linguístico e a inovação tecnológica para o avanço de sistemas de IA aplicados à fala. O vínculo com o ODS 10 (“Redução das Desigualdades”) se expressa na promoção da acessibilidade comunicativa, que amplia a inclusão de pessoas com deficiência visual, dislexia ou outras dificuldades associadas ao processo de leitura e à fluência prosódica. Por fim, o compromisso com o ODS 17 (“Parcerias e Meios de Implementação”) se concretiza na perspectiva de fomentar a cooperação entre as universidades e as empresas de tecnologia digital, voltada ao aprimoramento de soluções sustentáveis e socialmente relevantes.[112]

A empresa Google tem a possibilidade de se beneficiar dos resultados da pesquisa, visto que eles comprovam a consistência da entoação declarativa neutra no português brasileiro em diferentes vozes do sistema de síntese de fala descrito, assim como a manutenção do contorno entoacional independentemente do modelo empregado. Esses achados conferem à empresa uma vantagem competitiva no mercado de soluções de síntese de

[112] Deve-se ressaltar que a utilização de tecnologia digital em pesquisas de diferentes áreas, por si só, não é suficiente para garantir a inovação. O trabalho de Massini-Cagliari (2013), apesar de não empregar a tecnologia digital, é inovador ao promover avanços nos estudos medievais por meio da análise dos aspectos sonoros do Português Arcaico. Esse trabalho, de suma relevância acadêmica, demonstra que a inovação não decorre da aplicação de novas tecnologias, mas da investigação de assuntos ainda pouco examinados ou que necessitam de esclarecimentos, a partir de uma pertinente fundamentação teórica e de uma rigorosa metodologia.

fala. Além da verificação da eficácia da tecnologia, os resultados salientam a existência de controles prosódicos sofisticados, de mecanismos automáticos que asseguram a sinalização de fronteiras melódicas, mesmo em contextos com a ausência de material sonoro por razões linguísticas, e de métricas prosódicas aprimoradas, que integram os processos de avaliação da qualidade da estrutura entoacional dos enunciados declarativos neutros. Tais características podem tornar a fala sintética mais natural, favorecer a aceitação social do sistema e consolidar a aplicação da tecnologia em situações inclusivas e educacionais, além de estimular a colaboração entre os linguistas e os profissionais das áreas computacionais, eletrônicas e estatísticas.[113]

Em suma, a pesquisa aborda dois aspectos fundamentais. O primeiro investiga a entoação do português brasileiro na modalidade declarativa neutra e amplia o conhecimento científico referente ao tema. O segundo, por sua vez, analisa o papel da tecnologia de síntese de fala em aplicações linguístico-computacionais, com ênfase na capacidade do Google Cloud Text-to-Speech de reproduzir os padrões melódicos da variedade brasileira do português em enunciados declarativos neutros. Essas perspectivas acadêmicas comprovam que a investigação da prosódia sintética tem um vasto e promissor escopo, tanto para o trabalho linguístico quanto para o desenvolvimento digital, além de estimular o fortalecimento das iniciativas empresariais e das linhas de pesquisa na interface entre a linguagem e a IA.

[113] O trabalho aborda apenas um sistema de síntese de fala, mas os resultados obtidos são aplicáveis ao desenvolvimento de outras tecnologias equivalentes. A pesquisa destaca a importância de representações entoacionais para o treinamento e o controle de métricas de densidade tonal e de amplitude melódica, além da utilização de mecanismos fonéticos que asseguram a realização de fronteiras entoacionais e de uma avaliação que combine a inspeção auditiva, a análise acústica e, em etapas futuras, os experimentos psicoacústicos. Esses elementos definem um conjunto de diretrizes sólidas para a construção e a análise de módulos prosódicos em atuais e futuros sistemas de síntese de fala baseados em IA.

REFERÊNCIAS

ABAURRE-GNERRE, Maria Bernadete Marques. Processos fonológicos segmentais como índices de padrões prosódicos diversos nos estilos formal e casual do português do Brasil. **Caderno de Estudos Lingüísticos**, Campinas, v. 2, p. 23-44, 1981.

ABAURRE, Maria Bernadete Marques. Fonologia: a gramática dos sons. **Letras**, Santa Maria, n. 5, p. 9-24, jan./jun. 1993.

ABAURRE, Maria Bernadete Marques. Acento frasal e os processos fonológicos segmentais. **Letras de Hoje**, Porto Alegre, v. 31, n. 2, p. 41-50, 1996.

ABAURRE, Maria Bernadete Marques. Contribuições da física estatística e do formalismo termodinâmico para a modelagem da aquisição e mudança lingüística. **Intercâmbio**, São Paulo, v. 7, p. 187-199, 1998.

ABAURRE, Maria Bernadete Marques. Horizontes e limites de um programa de investigação em aquisição da escrita. *In*: LAMPRECHT, Regina (org.). **Aquisição da linguagem**: questões e análises. Porto Alegre: EDIPUCRS, 1999. v. 1, p. 167-186.

ABAURRE, Maria Bernadete Marques. A fonologia na gramática do português falado. **Letras de Hoje**, Porto Alegre, v. 38, n. 4, p. 35-48, 2003.

ABAURRE, Maria Bernadete Marques. Fonologia e Fonética. *In*: GUIMARÃES, Eduardo; ZOPPI-FONTANA, Mónica (org.). **Introdução às ciências da linguagem**: a palavra e a frase. 2. ed. Campinas: Pontes Editores, 2010. p. 39-74.

ABAURRE, Maria Bernadete Marques. Apresentação. *In*: ABAURRE, Maria Bernadete Marques (org.). **Gramática do português culto falado no Brasil**. Volume VII: a construção fonológica da palavra. São Paulo: Contexto, 2013. p. 9-18.

ABAURRE, Maria Bernadete Marques. Monotongações e ditongações. *In*: HORA, Dermeval da; BATTISTI, Elisa; MONARETTO, Valéria Oliveira (org.). **História do português brasileiro**. Volume III: mudança fônica do português brasileiro. São Paulo: Contexto, 2019. p. 78-107.

ABAURRE, Maria Bernadete Marques; WETZELS, Leo. Sobre a estrutura da gramática fonológica. **Cadernos de Estudos Lingüísticos**, Campinas, v. 23, p. 5-18, 1992.

ABBASS, Hussein. What is Artificial Intelligence? **IEEE Transactions on Artificial Intelligence**, Piscataway, v. 2, n. 2, p. 94-95, 2021.

ABERCROMBIE, David. **Elements of General Phonetics**. Edinburgh: Edinburgh University Press, 1967.

ALBANO, Eleonora Cavalcante. Fonologia de laboratório. *In*: HORA, Dermeval da; MATZENAUER, Carmen Lúcia (org.). **Fonologia, fonologias**: uma introdução. São Paulo: Contexto, 2017. p. 169-182.

ALBANO, Eleonora Cavalcante; MOREIRA, Agnaldo; SILVA, Adelaide Hercília Pescatori;

AQUINO, Patrícia Aparecida de; KAKINOHANA, Régis. Um conversor ortográfico-fônico e uma notação prosódica mínima para síntese de fala em língua portuguesa. *In*: SCARPA, Ester Mirian (org.). **Estudos de Prosódia**. Campinas: Editora da Unicamp, 1999. p. 85-105.

AQUINO, Patrícia Aparecida de. **O papel das vogais reduzidas pós-tônicas na construção de um sistema de síntese concatenativa para o português do Brasil**. 1997. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1997.

BAE, Jae-Sung; BAE, Hanbin; JOO, Young-Sun; LEE, Junmo; LEE, Gyeong-Hoon; CHO, Hoon-Young. Speaking speed control of end-to-end speech synthesis using sentence-level conditioning. *In*: INTERSPEECH, 2020, Shanghai. **Proceedings** [...]. Shanghai: ISCA, 2020. p. 4402-4406.

BARBOSA, Plinio Almeida. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis onautomatic segmental duration generation. **Cadernos de Estudos Linguísticos**, Campinas, v. 31, p. 33-53, 1996.

BARBOSA, Plinio Almeida. Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração de ciência e tecnologia de fala. *In*: SCARPA, Ester Mirian (org.). **Estudos de Prosódia**. Campinas: Editora da Unicamp, 1999. p. 21-52.

BARBOSA, Plinio Almeida. Máquinas falantes como instrumentos lingüísticos: por um humanismo éclairé. **Línguas e Instrumentos Lingüísticos**, Campinas, v. 8, p. 51-99, 2001.

BARBOSA, Plinio Almeida. **Incursões em torno do ritmo da fala**. Campinas: Pontes, 2006.

BARBOSA, Plinio Almeida. **Prosódia**. São Paulo: Parábola, 2019.

BARBOSA, Plinio Almeida. **As ciências da fala**. São Paulo: Parábola, 2022.

BARBOSA, Plinio Almeida; VIOLARO, Fábio; ALBANO, Eleonora Cavalcante; SIMÕES, Flávio; AQUINO, Patrícia Aparecida de; MADUREIRA, Sandra; FRANÇOZO, Edson. Aiuruete: a high-quality concatenative text-to-speech system for Brazilian Portuguese with demisyllabic analysis-based units and a hierarchical model of rhythm production. *In*: EUROSPEECH – 6th EUROPEAN CONFERENCE ON SPEECH COMMUNICATION AND TECHNOLOGY, 1999, Budapeste. **Proceedings** […]. Budapeste: ISCA, 1999. p. 2059-2062.

BARBOSA, Plinio Almeida; SILVA, Wellington da. A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches. *In*: International Conference on Computational Processing of the Portuguese Language, 10, 2012, Coimbra. **Proceedings** [...]. Heidelberg: Springer, 2012. p. 329-337. Lecture Notes in Computer Science (LNCS, v. 7243).

BARROS, Maria João Almeida de Sá. **Estudo comparativo e técnicas de geração de sinal para a síntese da fala**. 2002. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) – Faculdade de Engenharia, Universidade do Porto, Porto, 2002.

BECKMAN, Mary Esther; PIERREHUMBERT, Janet Breckenridge. Intonational Structure in

Japanese and English. **Phonology Yearbook**, Cambridge, n. 3, p. 255-310, 1986.

BEERENDS, John G. The influence of duration on the perception of pitch in single and simultaneous complex tones. **J. Acoust. Soc. Am.**, Melville, v. 86, p. 1835-1844, 1989.

BERTI, Larissa Cristina. Relação entre produção e percepção de fala: coerência com o parâmetro fonético-acústico. **Cadernos de Estudos Linguísticos**, Campinas, v. 50, p. 45-67, 2008.

BERTI, Larissa Cristina. PERCEFAL: an instrument to assess identification of phonological contrasts in Brazilian Portuguese. **Audiology-Communication Research**, São Paulo, v. 22, p. 1-9, 2017.

BERTI, Larissa Cristina; SPAZZAPAN, Evelyn Alves; QUEIROZ, Marcelo; PEREIRA, Pedro Leyton; FERNANDES-SVARTMAN, Flaviane Romani; RAPOSO DE MEDEIROS, Beatriz; MARTINS, Marcus Vinıcius Moreira; FERREIRA, Letıcia Santiago; DA SILVA, Ingrid Gandolfi Gomes; SABINO, Ester Cerdeira; LEVIN, Anna Sara; FINGER, Marcelo. Fundamental frequency related parameters in Brazilians with COVID-19. **J. Acoust. Soc. Am.**, Melville, v. 153, p. 576-585, 2023.

BERTI, Larissa Cristina; GAUY, Marcelo; DA SILVA, Luana Cristina Santos; RIOS, Julia Vasquez Valenci; MORAIS, Viviam Batista; ALMEIDA, Tatiane Cristina de; SOSOLETE, Leisi Silva; QUIRINO, José Henrique de Moura; MARTINS, Carolina Fernanda Pentean; FERNANDES-SVARTMAN, Flaviane Romani; RAPOSO DE MEDEIROS, Beatriz; QUEIROZ, Marcelo; GAZZOLA, Murilo; FINGER, Marcelo. Acoustic characteristics of voice and speech in post-COVID-19. **Healthcare**, Basel, v. 13, n. 1, p. 63, 2025.

BISOL, Leda. Mattoso Câmara Jr. e a palavra prosódica. **DELTA**: Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 20, n. Edição Especial, p. 59-70, 2004.

BISOL, Leda. Os constituintes prosódicos. *In*: BISOL, Leda (org.). **Introdução a estudos de fonologia do português brasileiro**. 5. ed. rev. Porto Alegre: EDIPUCRS, 2014. p. 259-271.

BOLINGER, Dwight (ed.). **Intonation**. London: Harmondsworth, 1972.

BOLINGER, Dwight. **Intonation and its parts**: Melody in spoken English. Palo Alto: Stanford University Press, 1986.

BRIGNER, Willard L. Perceived duration as a function of pitch. **Perceptual and Motor Skills**, Thousand Oaks, v. 67, n. 1, p. 301-302, 1988.

BOERSMA, Paul; WEENINK, David. **Manual**: the documentation system for the Praat program. Amsterdam: Institute of Phonetic Sciences, University of Amsterdam, [*s.d.*]. Disponível em: <https://www.praat.org/manual/Manual.html>. Acesso em: 11 de novembro de 2025.

BOERSMA, Paul; WEENINK, David. **Praat**: doing phonetics by computer. Versão 6.4.04. [Programa de computador], 2024. Disponível em: <http://www.praat.org/>. Acesso em: 20 de janeiro de 2024.

CAGLIARI, Luiz Carlos. **Elementos de Fonética do Português Brasileiro**. Tese (Livre Docência) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1981.

CAGLIARI, Luiz Carlos. Análise fonética do ritmo em poesia. **EPA**: estudos portugueses e africanos, Campinas, n. 3, p. 67-96, 1984.

CAGLIARI, Luiz Carlos. O ritmo do português na interpretação de Jerônimo Soares Barbosa. *In*: I ENCONTRO NACIONAL DE FONÉTICA E FONOLOGIA, 1985, Florianópolis. **Anais** [...]. Florianópolis: Universidade Federal de Santa Catarina, 1985. p. 27-38.

CAGLIARI, Luiz Carlos. Marcadores prosódicos na escrita. *In*: SEMINÁRIO DO GRUPO DE ESTUDOS LINGUÍSTICOS DO ESTADO DE SÃO PAULO, 18, 1989, Lorena. **Anais** [...]. Lorena: Grupo de Estudos Linguísticos de São Paulo, 1989. p. 195-203.

CAGLIARI, Luiz Carlos. Prosódia: algumas funções dos supra-segmentos. **Cadernos de Estudos Linguísticos**, Campinas, v. 23, p. 137-151, jul./dez. 1992.

CAGLIARI, Luiz Carlos. Da importância da prosódia na descrição de fatos gramaticais. *In*: ILARI, Rodolfo (org.). **Gramática do português falado**. Volume II: níveis de análise linguística. 4. ed. rev. Campinas: Editora da UNICAMP, 2002a. p. 37-60.

CAGLIARI, Luiz Carlos. **Análise fonológica**: introdução à teoria e à prática com especial destaque para o modelo fonêmico. Campinas: Mercado de Letras, 2002b.

CAGLIARI, Luiz Carlos. **Elementos de Fonética do Português Brasileiro**. São Paulo: Paulistana, 2007.

CAGLIARI, Luiz Carlos. Entoação e Fonologia. **Estudos Linguísticos (São Paulo. 1978)**, v. 41, n. 1, p. 8-22, 2012a.

CAGLIARI, Luiz Carlos. Línguas de ritmo silábico. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 20, n. 2, p. 23-58, 2012b.

CAGLIARI, Luiz Carlos; ABAURRE, Maria Bernadete Marques. Elementos para uma investigação instrumental das relações entre padrões rítmicos e processos fonológicos no português brasileiro. **Cadernos de Estudos Linguísticos**, Campinas, v. 10, p. 39-57, 1986.

CAGLIARI, Luiz Carlos; MASSINI-CAGLIARI, Gladis. O papel da tessitura dentro da prosódia portuguesa. *In*: CASTRO, Ivo; DUARTE, Inês (org.). **Razões e emoção**: miscelânea de estudos em homenagem a Maria Helena Mira Mateus. Lisboa: Imprensa Nacional - Casa da Moeda, 2003. v. 1., p. 67-85.

CALLOU, Dinah; LEITE, Yonne. **Iniciação à fonética e à fonologia do português**. 3. ed. rev. Rio de Janeiro: Zahar, 1994.

CALLOU, Dinah; SERRA, Carolina Ribeiro. Variação do rótico e estrutura prosódica. **Revista do GELNE**, Natal, v. 14, n. 1/2, p. 41-57, 2016.

CAMARA JR., Joaquim Mattoso. **Problemas de lingüística descritiva**. Petrópolis: Vozes,

1969.

CAMARA JR., Joaquim Mattoso. **História e estrutura da língua portuguesa**. Rio de Janeiro: Padrão, 1975.

CAMARA JR., Joaquim Mattoso. **Estrutura da Língua Portuguesa**. 47. ed. Petrópolis: Vozes, 2015.

CASANOVA, Edresson. **Síntese de voz aplicada ao português brasileiro usando aprendizado profundo**. 2019. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Tecnológica Federal do Paraná, Medianeira, 2019.

CASANOVA, Edresson. **Síntese de fala aplicada à geração de conjunto de dados para reconhecimento automático de fala**. 2022. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022.

CASANOVA, Edresson; SHULBY, Christopher; ALUÍSIO, Sandra Maria. Deep Learning approaches for Speech Synthesis and Speaker Verification. *In*: OTTA, Emma; MONTICELLI, Patrícia Ferreira (org.). **Acoustic communication**: an interdisciplinary approach. São Paulo: Instituto de Psicologia da Universidade de São Paulo, 2021.

CASANOVA, Edresson; SHULBY, Christopher; GÖLGE, Eren; MÜLLER, Nicolas Michael; OLIVEIRA, Frederico Santos de; CANDIDO JR., Arnaldo; SOARES, Anderson da Silva; ALUÍSIO, Sandra Maria; PONTI, Moacir Antonelli. SC-GlowTTS: an Efficient Zero-Shot Multi-Speaker Text-To-Speech Model. *In*: INTERSPEECH, 2021, Brno. **Proceedings** [...]. Brno: ISCA, 2021. p. 3645-3649.

CASANOVA, Edresson; WEBER, Julian; SHULBY, Christopher; CANDIDO JR., Arnaldo; GÖLGE, Eren; PONTI, Moacir Antonelli. YourTTS: Towards Zero-Shot Multi-Speaker TTS and Zero-Shot Voice Conversion for Everyone. *In*: 39TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 2022, Baltimore. **Proceedings** [...]. Baltimore: PMLR, v. 162, 2022. p. 2709-2720.

CASANOVA, Edresson; DAVIS, Kelly; GÖLGE, Eren; GÖKNAR, Görkem; GULEA, Iulian; HART, Logan; ALJAFARI, Aya; MEYER, Joshua; MORAIS, Reuben; OLAYEMI, Samuel; WEBER, Julian. XTTS: a Massively Multilingual Zero-Shot Text-to-Speech Model. *In*: INTERSPEECH, 2024, Kos. **Proceedings** [...]. Kos: ISCA, 2024. p. 4978-4982.

CASTELO, Joelma. **A entoação dos enunciados declarativos e interrogativos no português do Brasil:** uma análise fonológica em variedades ao longo da Costa Atlântica. 2016. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade de Lisboa, Lisboa, 2016.

CASTILHO, Ataliba Teixeira. **Nova gramática do português brasileiro**. São Paulo: Contexto, 2019.

CAVALCANTE, Marianne Carvalho Bezerra; SCARPA, Ester Mirian. Prosódia e aquisição da linguagem. *In*: OLIVEIRA JR., Miguel (org.). **Prosódia, prosódias**: uma introdução. São Paulo: Contexto, 2022. v. 1, p. 97-110.

CHOMSKY, Noam. **Syntactic Structures**. Haia: Mouton, 1957.

CHOMSKY, Noam. **Aspects of the theory of syntax**. Cambridge, MA: The MIT Press, 1965.

CHOMSKY, Noam. **Knowledge of language**: its nature, origin, and use. New York: Praeger, 1986.

CHOMSKY, Noam. The minimalist program for linguistic theory. *In*: HALE, Kenneth; KEYSER, Samuel Jay (org.). **The view from Building 20**: essays in linguistics in honor of Sylvain Bromberger. Cambridge, MA: The MIT Press, 1993. p. 1-52.

CHOMSKY, Noam. **O programa minimalista**. Trad. Eduardo Raposo. Lisboa: Caminho, 1999.

CHOMSKY, Noam; HALLE, Morris. **The Sound Pattern of English**. New York: Harper & Row, 1968.

COLLISCHON, Gisela. A sílaba em Português. *In*: BISOL, Leda (org.). **Introdução a Estudos de Fonologia do Português Brasileiro**. Porto Alegre: EDIPUCRS, 2005. p. 101-130.

COLLISCHONN, Gisela. Traçando percursos da fonologia. **Revista da Anpoll**, Florianópolis, v. 1, n. 29, 2010.

CONSTANTINI, Ana Carolina. **Caracterização prosódica de sujeitos de diferentes variedades de fala do português brasileiro em diferentes relações sinal-ruído**. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2014.

CONSTANTINI, Ana Carolina; BARBOSA, Plinio Almeida. Prosodic characteristics of different varieties of Brazilian Portuguese. **Revista Brasileira de Criminalística**, Brasília, v. 4, n. 3, p. 44-53, 2015.

COOPER, Franklin S.; LIBERMAN, Alvin M.; BORST, John M.; GERSTMAN, H. The interconversion of audible and visible patterns as a basis for research in the perception of speech. **Proc. Natl. Acad. Sci.**, Washington, D.C., v. 37, p. 318-325, 1951.

CÓRDULA, Maíra Sueco Maegava. **Análise fonético-fonológica dos padrões entoacionais do português brasileiro e do inglês norte-americano no filme *Shrek* (2001)**. 2012. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, Araraquara, 2012.

CÓRDULA, Maíra Sueco Maegava. **Entoação e sentidos**: análise fonético-fonológica dos padrões entoacionais do português brasileiro e do inglês norte-americano no filme *Shrek* (2001). São Paulo: Cultura Acadêmica, 2013.

COUPER-KUHLEN, Elizabeth. **An introduction to English prosody**. London: Edward Arnold, 1986.

CRISTÓFARO SILVA, Thaïs. **Dicionário de Fonética e Fonologia**. São Paulo: Contexto,

2015.

CRISTÓFARO SILVA, Thaïs. Fonologia: contribuições para a Linguística e para a Computação. **Estudos Linguísticos (São Paulo. 1978)**, São Paulo, v. 40, n. 1, p. 33-46, jan./abr. 2011.

CRISTÓFARO SILVA, Thaïs. **Fonética e fonologia do português**: roteiro de estudos e guia de exercícios. 11. ed. São Paulo: Contexto, 2022.

CRISTÓFARO SILVA, Thaïs. **Instruções básicas para usar o Praat**. Fonologia.org, 2021a. Disponível em: <https://fonologia.org/fonetica-acustica-o-som/>. Acesso em: 18 de março de 2024.

CRISTÓFARO SILVA, Thaïs. **Instruções básicas para anotar e segmentar dados no Praat**. Fonologia.org, 2021b. Disponível em: <https://fonologia.org/fonetica-acustica-o- som/>. Acesso em: 18 de março de 2024.

CRISTÓFARO SILVA, Thaïs; SEARA, Izabel Christine; SILVA, Adelaide Hercília Pescatori; RAUBER, Andreia Schurt; CANTONI, Maria Mendes. **Fonética Acústica**: os sons do português brasileiro. São Paulo: Contexto, 2019.

CRUTTENDEN, Alan. **Intonation**. Cambridge: Cambridge University Press, 1986.

CRUZ, Marisa; FROTA, Sónia. O sintagma entoacional na gaguez: evidências do PE. *In*: BRITO, Ana Maria; SILVA, Fátima; VELOSO, João; FIÉIS, Alexandra (ed.). **XXV Encontro Nacional da Associação Portuguesa de Linguística**: textos seleccionados. Porto: Associação Portuguesa de Linguística, 2010. p. 365-383.

CUMMING, Ruth E. **Speech rhythm**: The language-specific integration of pitch and duration. 2010. Tese (Doutorado) – University of Cambridge, Cambridge, 2010.

CUNHA, Cláudia de Souza. **Entoação regional no Português do Brasil**. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.

DUTOIT, Thierry. **An introduction to text-to-speech synthesis**. Norwell: Kluwer Academic Publishers, 1997.

EGASHIRA, Francisco. **Síntese de voz a partir de texto para a língua portuguesa**. 1992. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica, Campinas, 1992.

EISENSTEIN, Jacob. **Introduction to Natural Language Processing**. Cambridge, MA; London, England: The MIT Press, 2019.

ESCOVEDO, Tatiana; KOSHIYAMA, Adriano. **Introdução a Data Science**: algoritmos de Machine Learning e métodos de análise. São Paulo: Casa do Código, 2020.

FERNANDES, Flaviane Romani. Tonal association in neutral and subject-narrow-focus sentences of Brazilian Portuguese: a comparison with European Portuguese. **Journal of**

Portuguese Linguistics, Lisboa, v. 6, n. 1, p. 91-115, 2007a.

FERNANDES, Flaviane Romani. **Ordem, focalização e preenchimento em português**: sintaxe e prosódia. 2007b. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2007b.

FERNANDES-SVARTMAN, Flaviane Romani. Acento secundário, atribuição tonal e ênfase em português brasileiro (PB). **Estudos Lingüísticos**, São Paulo, v. 38, n. 1, p. 47-58, 2009.

FERNANDES-SVARTMAN, Flaviane Romani. A entoação das sentenças clivadas em português brasileiro e a interface sintaxe-fonologia. **Filologia e Linguística Portuguesa**, São Paulo, v. 14, n. 1, p. 37-56, 2012.

FERNANDES-SVARTMAN, Flaviane Romani. **Fraseamento entoacional e densidade tonal em português brasileiro**: elementos para a discussão da relação entre constituintes entoacionais e domínios prosódicos. 2024a. Tese (Livre-Docência) – Universidade de São Paulo, São Paulo, 2024a.

FERNANDES-SVARTMAN, Flaviane Romani. Current issues in Brazilian Portuguese intonation. *In*: ZAMPAULO, André (ed.). **The Routledge Handbook of Portuguese Phonology**. London: Routledge, 2024b. p. 141-163.

FERNANDES-SVARTMAN, Flaviane Romani; ROMANO, Nícolas. Fatores determinantes na associação tonal em sentenças neutras do português brasileiro. **Cadernos de Estudos Linguísticos**, Campinas, v. 59, n. 3, p. 537-553, 2017.

FERREIRA, Marcelo; LOPES, Marcos. Linguística Computacional. *In*: FIORIN, José Luiz (org.). **Novos Caminhos da Linguística**. São Paulo: Contexto, 2017.

FERREIRA, Marcelo; LOPES, Marcos. **Para conhecer Linguística Computacional**. São Paulo: Contexto, 2021.

FÓNAGY, Ivan. As funções modais da entoação. **Cadernos de Estudos Linguísticos**, Campinas, n. 25, p. 25-65, 1993.

FREITAS, Cláudia. **Linguística Computacional**. São Paulo: Parábola, 2022.

FROTA, Sónia. **Prosody and focus in European Portuguese**: phonological phrasing and intonation. New York: Garland Publishing, 2000.

FROTA, Sónia; VIGÁRIO, Marina. Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. *In*: CASTRO, Rui Vieira de; BARBOSA, Pilar (org.). **Actas do XV Encontro Nacional da Associação Portuguesa de Linguística**. Coimbra: APL, 2000, v. 1, p. 533-555.

FROTA, Sónia; D'IMPERIO, Mariapaola; ELORDIETA, Gorka; PRIETO, Pilar; VIGÁRIO, Marina. The phonetics and phonology of intonational phrasing in Romance. *In*: PRIETO, Pilar; MASCARÓ, Joan; SOLÉ, Maria-Josep (ed.). **Segmental and Prosodic Issues in Romance Phonology**. Amsterdam: John Benjamins, 2007. p. 131-153.

FROTA, Sónia; BUTLER, Joseph; VIGÁRIO, Marina. Infants' perception of intonation: is it a

statement or a question? **Infancy**, Hoboken, v. 19, n. 2, p. 194-213, 2014.

FROTA, Sónia; OLIVEIRA, Pedro; CRUZ, Marisa; VIGÁRIO, Marina. **P-ToBI**: tools for the transcription of Portuguese prosody. Lisboa: Laboratório de Fonética, CLUL/FLUL, 2015.

FROTA, Sónia; CRUZ, Marisa; FERNANDES-SVARTMAN, Flaviane Romani; COLLISCHONN, Gisela; FONSECA, Aline; SERRA, Carolina; OLIVEIRA, Pedro; VIGÁRIO, Marina. Intonational variation in Portuguese: European and Brazilian varieties. *In*: FROTA, Sónia; PRIETO, Pilar (ed.). **Intonation in Romance**. Oxford: Oxford University Press, 2015. p. 235-283.

FROTA, Sónia; MORAES, João Antônio de. Intonation in European and Brazilian Portuguese. *In*: WETZELS, Willem Leo; COSTA, João; MENUZZI, Sérgio (ed.). **The Handbook of Portuguese Linguistics**. Chichester: John Wiley & Sons, Inc, 2016, p. 141-166.

GALDINO, Julio Cesar. **Em 200 metros, vire à esquerda**: a entoação dos comandos de GPS. 2023. Dissertação (Mestrado em Linguística) – Universidade Federal de Alagoas, Faculdade de Letras, Maceió, 2023.

GALDINO, Julio Cesar; SILVA, Kyvia Fernanda Tenório da; OLIVEIRA JR., Miguel. Características prosódicas associadas aos sinais de pontuação: uma revisão de escopo. **Cadernos de Linguística**, Campinas, v. 2, n. 4, p. e468, 2021.

GALDINO, Julio Cesar; OLIVEIRA JR., Miguel. Prosódia e síntese da fala: uma revisão integrativa da literatura. **Revista da ABRALIN**, Campinas, v. 22, n. 1, p. 1-15, 2023.

GALVÃO PASSETTI, Gabriel Henrique. **Coordenação de constituintes não oracionais por meio de *mas* nas variedades portuguesas sob a perspectiva da Gramática Discursivo-Funcional**: Concessão e Contraste. 2021. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2021.

GALVES, Charlotte Marie Chambelland; ABAURRE, Maria Bernadete Marques. Os clíticos no português brasileiro: elementos para uma abordagem sintático-fonológica. *In*: CASTILHO, Ataliba Teixeira de; BASÍLIO, Margarida (org.). **Gramática do português falado**. Volume IV: estudos descritivos. 2. ed. rev. Campinas: Editora da UNICAMP, 2002. p. 267-312.

GOMES, Leandro de Campos Teixeira. **Sistema de conversão texto-fala para a língua portuguesa utilizando a abordagem de síntese por regras**. 1998. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, 1998.

GONÇALVES, Carlos Alexandre. Foco e topicalização: delimitação e confronto de estruturas. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 7, n. 1, p. 31–50, jan./jun. 1998.

GOOGLE CLOUD. **Google Cloud Text-to-Speech API**. Disponível em: <https://cloud.google.com/text-to-speech>. Acesso em: 2 de março de 2024.

GOOGLE CLOUD. **Text-to-Speech**. Disponível em: <https://cloud.google.com/text-to-speech?hl=pt-br>. Acesso em: 24 de julho de 2023.

GOOGLE CLOUD. **WaveNet voices**. Disponível em: <https://cloud.google.com/text-to-speech/docs/wavenet?hl=pt-br>. Acesso em: 8 de setembro de 2023.

GOOGLE CLOUD. **Text-to-Speech**: benefits. Disponível em: <https://cloud.google.com/text-to-speech#benefits>. Acesso em: 13 de julho de 2024.

GOOGLE CLOUD. **Text-to-Speech**. Disponível em: <https://cloud.google.com/text-to-speech?hl=pt_br>. Acesso em: 30 de dezembro de 2024.

GOOGLE CLOUD. **Text-to-Speech**. Disponível em: <https://docs.cloud.google.com/text-to-speech/docs/list-voices-and-types?hl=pt-br#list_of_all_supported_languages>. Acesso em: 31 de outubro de 2025.

GOMES DA SILVA, Carolina; MIRANDA, Luma da Silva; CARNAVAL, Manuella; CUNHA, Cláudia de Souza. A entoação da ordem no Português do Brasil: uma descrição dialetal a partir do *corpus* ALiB. **Journal of Speech**, Campinas, v. 5, n. 2, p. 29-45, 2016.

GUSSENHOVEN, Carlos. Phonology of intonation. **Glot International**, Oxford, v. 6, n. 9/10, p. 271-284, 2002.

HALLIDAY, Michael Alexander Kirkwood. The tones of English. *In*: JONES, W. E.; LAVER, John (ed.). **Phonetics in Linguistics**: a book of readings. London: Longman, 1963. p. 103-126.

HALLIDAY, Michael Alexander Kirkwood. **A Course in spoken English**: intonation. London: Oxford University Press, 1970.

HAYES, Bruce; LAHIRI, Aditi. Bengali intonational phonology. **Natural Language & Linguistic Theory**, Dordrecht, v. 9, n. 1, p. 47-96, 1991.

JUN, Sun-Ah. **The Phonetics and Phonology of Korean Prosody**: Intonational Phonology and Prosodic Structure. Nova York: Garland Publishing Inc., 1996.

KALINOWSKI, Marcos; ESCOVEDO, Tatiana; VILLAMIZAR, Hugo; LOPES, Hélio. **Engenharia de Software para Ciência de Dados**: um guia de boas práticas com ênfase na construção de sistemas de Machine Learning em Python. São Paulo: Casa do Código, 2023.

KATO, Mary Aizawa. Sintaxe e aquisição na Teoria de Princípios e Parâmetros. **Letras de Hoje**, Porto Alegre, v. 30, n. 4, p. 57-73, dez. 1995.

KATO, Mary Aizawa. Teoria sintática: de uma perspectiva de "-ismos" para uma perspectiva de "programas". **DELTA**: Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 13, n. 2, p. 5-22, ago. 1997.

KATO, Mary Aizawa. Aquisição da linguagem numa abordagem gerativista. **Letras de Hoje**, Porto Alegre, v. 34, n. 3, p. 17-25, set. 1999.

KATO, Mary Aizawa. Nomes e pronomes na aquisição. **Letras de Hoje**, Porto Alegre, v. 36, n. 3, p. 101-112, set. 2001.

KATO, Mary Aizawa. A evolução da noção de parâmetros. **DELTA**: Documentação de Estudos em Linguística Teórica e Aplicada, São Paulo, v. 18, n. 2, p. 309-338, 2002.

KATO, Mary Aizawa. A contribuição chomskiana para a compreensão da aprendizagem de L2. **Trabalhos em Linguística Aplicada**, Campinas, v. 44, n. 2, p. 185-199, jul./dez. 2005.

KATO, Mary Aizawa; NASCIMENTO, Milton do. Apresentação. *In*: KATO, Mary Aizawa; NASCIMENTO, Milton do (org.). **Gramática do português culto falado no Brasil**. Volume II: a construção da sentença. São Paulo: Contexto, 2020. p. 11-18.

KATO, Mary Aizawa; MIOTO, Carlos. A arquitetura da gramática. *In*: KATO, Mary Aizawa; NASCIMENTO, Milton do (org.). **Gramática do português culto falado no Brasil**. Volume II: a construção da sentença. São Paulo: Contexto, 2020. p. 19-36.

KEMPELEN, Wolfgang von. **Mechanismus der Menschlichen Sprache Nebst der Beschreibung Seiner Sprechenden Maschine**. Wien: Degen, 1791. Disponível em: <https://www.deutschestextarchiv.de/book/show/kempelen_maschine_1791> Acesso em: 14 de abril de 2025.

KINGSTON, John; BECKMAN, Mary Esther (ed.). **Papers in Laboratory Phonology I**: Between the Grammar and Physics of Speech. Cambridge: Cambridge University Press, 1990.

LADD, Dwight Robert. **Intonational Phonology**. Cambridge: Cambridge University Press, 1996.

LADD, Dwight Robert. **Intonational Phonology**. 2. ed. Cambridge: Cambridge University Press, 2008.

LADEFOGED, Peter. **Elements of acoustic phonetics**. Chicago: University of Chicago Press, 1962.

LAVER, John. **Principles of phonetics**. Cambridge: Cambridge University Press, 1994.

LEE, Seung Hwa. Fonologia Gerativa. *In*: HORA, Dermeval da; MATZENAUER, Carmen Lúcia (org.). **Fonologia, fonologias**: uma introdução. São Paulo: Contexto, 2017. p. 31-45.

LEIBNIZ ASSOCIATION. **A máquina de falar "Kempelen"**. Google Arts & Culture, [20--]. Disponível em: <https://artsandculture.google.com/story/2QUB7hLe64FKJA>. Acesso em: 28 de março de 2025.

LIMONGI, Ricardo. The use of artificial intelligence in scientific research with integrity and ethics. **Future Studies Research Journal**: Trends and Strategies, São Paulo, v. 16, n. 1, p. 1-10, 2024.

LUCENTE, Luciana. **DaTo**: um sistema de notação entoacional do português brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea. 2008. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de

Campinas, Campinas, 2008.

LUCENTE, Luciana. **Aspectos dinâmicos da fala e da entoação do português brasileiro**. 2012. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2012.

LUCENTE, Luciana. Uma abordagem fonética na Fonologia Entoacional. **Fórum Linguístico**, Florianópolis, v. 11, n. 1, p. 79-95, jan./mar. 2014.

LUCENTE, Luciana. Introdução à análise entoacional. *In*: FREITAG, Raquel Meister Ko; LUCENTE, Luciana (org.). **Prosódia da fala**: pesquisa e ensino. São Paulo: Blucher, 2017. v. 1, p. 7-25.

MADUREIRA, Sandra. Entoação e síntese de fala: modelos e parâmetros. *In*: SCARPA, Ester Mirian (org.). **Estudos de Prosódia**. Campinas: Editora da Unicamp, 1999. p. 53-68.

MADUREIRA, Sandra. The investigation of speech expressivity. *In*: MELLO, Heliana; PANUNZI, Alessandro; RASO, Tommaso (ed.). **Pragmatics and Prosody**: illocution, modality, attitude, information patterning and speech annotation. Firenze: Firenze University Press, v. 1, p. 101-118, 2011.

MADUREIRA, Sandra. Intonation and variation: the multiplicity of forms and senses. **Dialectologia**: Revista Electrònica, Barcelona, Special Issue VI, p. 57-74, 2016.

MADUREIRA, Sandra; SILVA, Cairo Humberto da; AQUINO, Patrícia Aparecida de. Pitch Patterns and Duration: Analysis and Synthesis. In: XIII INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 1995, Estocolmo. **Proceedings** [...]. Estocolmo: ICPhS, 1995. p. 406-409.

MAIA, Eleonora Motta. **No reino da fala**: a linguagem e seus sons. São Paulo: Ática, 1985.

MASSINI, Gladis. **A duração no estudo do acento e do ritmo do português**. 1991. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1991.

MASSINI-CAGLIARI, Gladis. **Acento e ritmo**. São Paulo: Contexto, 1992.

MASSINI-CAGLIARI, Gladis. Sobre a natureza fonética do acento em Português. **D.E.L.T.A.**, São Paulo, v. 9, p. 195-216, 1993.

MASSINI-CAGLIARI, Gladis. **Cantigas de amigo**: do ritmo poético ao lingüístico. Um estudo do percurso histórico da acentuação em Português. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 1995.

MASSINI-CAGLIARI, Gladis. **Do poético ao lingüístico no ritmo dos trovadores**: três momentos da história do acento. Araraquara: FCL, Laboratório Editorial, UNESP. São Paulo: Cultura Acadêmica, 1999.

MASSINI-CAGLIARI, Gladis. O conceito de pé como unidade rítmica: trajetória. *In*: SCARPA, Ester Mirian (org.). **Estudos de prosódia**. Campinas: Editora da UNICAMP, 1999.

p. 113-139.

MASSINI-CAGLIARI, Gladis. Inovação científica em estudos medievais: descobrindo os sons do Português Arcaico. **Revista da ANPOLL**, Florianópolis, v. 1, n. 34, p. 17-50, 2013.

MASSINI-CAGLIARI, Gladis. **A música da fala dos trovadores**: desvendando a prosódia medieval. São Paulo: Editora Unesp Digital, 2015.

MASSINI-CAGLIARI, Gladis. Dialogando com a notação musical na busca de uma metodologia para investigação da entoação do português medieval: análise da *Cantiga de Santa Maria* 9. *In*: FERNANDES, Geraldo Augusto (org.). **Série Estudos Medievais 5**: abordagens interdiscursivas no contexto da cultura medieval. Fortaleza: GT de Estudos Medievais da ANPOLL, 2017. p. 24-37.

MASSINI-CAGLIARI, Gladis. Acentos em nomes. *In*: HORA, Dermeval da; BATTISTI, Elisa; MONARETTO, Valéria Oliveira (org.). **História do Português Brasileiro**: mudança fônica do português brasileiro. São Paulo: Contexto, 2019. v. 3. p. 198-225.

MASSINI-CAGLIARI, Gladis. Em busca da prosódia inaudível: a duração musical como pista da constituição das frases entoacionais e dos enunciados prosódicos nas *Cantigas de Santa Maria*. **Cadernos de Linguística**, Campinas, v. 2, p. 1-20, 2021.

MASSINI-CAGLIARI, Gladis. Análise da letra e da música de *Cantigas de Santa Maria*: em busca de pistas dos limites do sintagma entoacional no Português Medieval. *In*: HORA, Dermeval da; HELMER, Ángela Helmer (org.). **Interseções linguísticas**: estudos diversos. São Paulo: Líquido Editorial, 2023. v. 1., p. 30-53.

MASSINI-CAGLIARI, Gladis; CAGLIARI, Luiz Carlos. Fonética. *In*: MUSSALIM, Fernanda; BENTES, Anna Christina (org.). **Introdução à lingüística**: domínios e fronteiras. São Paulo: Cortez, 2001. v. 1, p. 105-146.

MASSINI-CAGLIARI, Gladis; CAGLIARI, Luiz Carlos. Fonética. *In*: MUSSALIM, Fernanda; BENTES, Anna Christina (org.). **Introdução à linguística**: domínios e fronteiras. 9. ed. rev. São Paulo: Cortez, 2012. v. 1, p. 113-156.

MINEMATSU, Nobuaki; HASHIMOTO, Hiroya; HIRANO, Hiroko; SAITO, Daisuke. Development of a prosodic reading tutor of Japanese – effective use of TTS and F_0 contour modeling techniques for CALL. *In*: SLaTE 2015 – Speech and Language Technology in Education, 2015, Leipzig. **Proceedings** [...]. Leipzig: ISCA, 4-5 set. 2015. p. 189.

MIXDORFF, Hansjörg. Quantitative tone and intonation modeling across languages. *In*: INTERNATIONAL SYMPOSIUM ON TONAL ASPECTS OF LANGUAGES, with emphasis on tone languages (TAL), 2004, Beijing. **Proceedings** [...]. Beijing: ISCA, 2004. p. 137-142.

MONAGHAN, Alex I. C.; LADD, Dwight Robert. Symbolic output as the basis for evaluating intonation in text-to-speech systems. **Speech Communication**, v. 9, n. 4, p. 305-314, 1990.

MORAES, João Antônio de. The pitch accents in Brazilian Portuguese: analysis by synthesis.

In: SPEECH PROSODY, 2008, Campinas. **Proceedings** [...]. Campinas: ISCA, 2008. p. 389-398.

MORAES, João Antônio de. Fonética, fonologia e a entoação do português: a contribuição da fonologia experimental. **Diadorim**, Rio de Janeiro, v. 18, ed. especial, p. 8-30, 2016.

MORAES, João Antônio de; LEITE, Yonne. Ritmo e velocidade da fala na estratégia do discurso: uma proposta de trabalho. *In*: ILARI, Rodolfo (org.). **Gramática do Português Falado**. Volume II: níveis de análise linguística. Campinas: Editora da UNICAMP, 1992. p. 65-77.

MORAES, João Antônio de; RILLIARD, Albert. Entoação. *In*: OLIVEIRA JR., Miguel (org.). **Prosódia, prosódias**: uma introdução. São Paulo: Contexto, 2022, p. 45-66.

NAÇÕES UNIDAS NO BRASIL. **Objetivos de Desenvolvimento Sustentável (ODS)**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 18 de setembro de 2024.

NAÇÕES UNIDAS NO BRASIL. **Objetivos de Desenvolvimento Sustentável (ODS)**. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 24 de fevereiro de 2025.

NESPOR, Marina; VOGEL, Irene. Prosodic domains of external sandhi rules. *In*: VAN DER HULST, Harry; SMITH, Norval (ed.). **The structure of phonological representations**. v. 1. Dordrecht: Foris, 1982. p. 225-255.

NESPOR, Marina; VOGEL, Irene. **Prosodic Phonology**. Dordrecht: Foris Publications, 1986.

NESPOR, Marina; VOGEL, Irene. **Prosodic Phonology**: with a new foreword. Berlin: Mouton de Gruyter, 2007. Obra original publicada em 1986.

NOLAN, Francis. Intonational equivalence: an experimental evaluation of pitch scales. *In*: INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 15., 2003, Barcelona. **Proceedings** [...]. Barcelona: ICPhS, 2003. p. 771‑774.

OLIVEIRA, Leonardo Mendes de; VEIT, Eliane Angela; SCHNEIDER, Claudio. **Freqüência**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2002. Disponível em: <https://www.if.ufrgs.br/cref/ntef/som/freq.html>. Acesso em: 28 de março de 2025.

OLIVEIRA JR., Miguel. O que é entonação. *In*: OTHERO, Gabriel de Ávila; FLORES, Valdir do Nascimento (org.). **O que sabemos sobre a linguagem**: 51 perguntas e respostas sobre a linguagem humana. São Paulo: Parábola Editorial, 2022, p. 261-265.

OPENAI. **Data Analysis with ChatGPT**. Disponível em: <https://chatgpt.com/g/g-HMNcP6w7d-data-analyst>. Acesso em: 21 de abril de 2024.

OPENAI. **ChatGPT (GPT-4o mini)**: modelo de linguagem de inteligência artificial. Disponível em: <https://openai.com/>. Acesso em: 9 de agosto de 2024.

OPENAI. **ChatGPT (GPT-4o mini)**: modelo de linguagem de inteligência artificial. Disponível em: <https://openai.com/>. Acesso em: 10 de setembro de 2024.

OSTERMANN FILHO, Paulo Eduardo. **Desenvolvimento de regras de pronúncia para a síntese de fala em língua portuguesa**. 2002. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Instituto de Informática, Porto Alegre, 2002.

OTHERO, Gabriel de Ávila. Linguística computacional: uma breve introdução. **Letras de Hoje**, Porto Alegre, v. 41, n. 2, p. 341-351, jun. 2006.

PACHECO, Vera. **Estudo dos Marcadores Prosódicos através de uma investigação acústico-perceptual de textos lidos por falantes do português do Brasil**. 2003. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2003.

PAIVA, Vera Lúcia Menezes de Oliveira e. **Manual de pesquisa em estudos linguísticos**. São Paulo: Parábola, 2019.

PAIXÃO, Vivian Borges. **Tecnologia assistiva e fonologia do português do Brasil: aspectos prosódicos da fala sintetizada pelo *software* LianeTTS**. 2020. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020.

PEZATTI, Erotilde Goreti. **A ordem de palavras em português**: aspectos tipológicos e funcionais. 1992. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Universidade Estadual Paulista, Araraquara, 1992.

PEZATTI, Erotilde Goreti. **A ordem das palavras no português**. São Paulo: Parábola Editorial, 2014.

PIERREHUMBERT, Janet Breckenridge. **The phonology and phonetics of English intonation**. 1980. Tese (Doctor of Philosophy) – Massachusetts Institute of Technology. Massachusetts: M.I.T. Press, 1980.

PIERREHUMBERT, Janet Breckenridge. Synthesizing intonation. **J. Acoust. Soc. Am.**, Melville, v. 70, n. 4, p. 985-995, 1981.

PIERREHUMBERT, Janet Breckenridge. Prosody, Intonation, and Speech Technology. *In*: BATES, Madeleine; WEISCHEDEL, Ralph M. (ed.). **Challenges in Natural Language Processing**. Cambridge: Cambridge University Press, 1993. p. 257-282.

PIERREHUMBERT, Janet Breckenridge; BECKMAN, Mary Esther. **Japanese tone structure**. Cambridge: M.I.T. Press, 1988.

PIERREHUMBERT, Janet Breckenridge; BECKMAN, Mary Esther; LADD, Dwight Robert. Conceptual foundations of phonology as a laboratory science. *In*: BURTON-ROBERTS, Noel; CARR, Philip; DOCHERTY, Gerard (ed.). **Phonological Knowledge**: Conceptual and Empirical Issues. Oxford: Oxford University Press, 2000. p. 273-303.

PIKE, Kenneth L. **The intonation of American English**. Ann Arbor: The University of Michigan Press, 1945.

PITRELLI, John F.; BECKMAN, Mary Esther; HIRSCHBERG, Julia. Evaluation of prosodic transcription labeling reliability in the ToBI framework. *In*: 3RD INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 1994, Yokohama. **Proceedings** [...]. Yokohama: ICPhS/ISCA, 1994. v. 2, p. 123-126.

PORTELANCE, Eva; JASBI, Masoud. On the compatibility of generative AI and generative linguistics. **Nature Computational Science**, Nova Iorque, v. 5, p. 745-753, 2025.

QUENÉ, Hugo; KAGER, René. The derivation of prosody for text-to-speech from prosodic sentence structure. **Computer Speech and Language**, Estados Unidos, v. 6, n. 1, p. 77-98, 1992.

RUBIN, Philip; GOLDSTEIN, Louis. **The Pattern Playback**. New Haven: Haskins Laboratories, [20--]. Disponível em: <https://haskinslabs.org/research/features-and-demos/pattern-playback>. Acesso em: 28 de março de 2025.

SÁ, Felipe Cortez de. **Geração de prosódia para o português brasileiro em sistemas text-to-speech**. Monografia (Bacharelado em Ciência da Computação) – Centro de Ciências Exatas e da Terra, Universidade Federal do Rio Grande do Norte, Natal, 2018.

SANTOS, Vinícius Gonçalves; FERNANDES-SVARTMAN, Flaviane Romani. Padrões tonais nucleares de declarativas e interrogativas neutras do português angolano do Libolo. **Lingüística**, Montevideo, v. 36, n. 1, p. 33-52, jun. 2020.

SCARPA, Ester Mirian. Desenvolvimento da intonação e a organização da fala inicial. **Cadernos de Estudos Linguísticos**, Campinas, n. 14, p. 65-84, 1988.

SCARPA, Ester Mirian. Apresentação. *In*: SCARPA, Ester Mirian (org.). **Estudos de prosódia**. Campinas: Editora da UNICAMP, 1999. p. 7-17.

SELKIRK, Elisabeth. **Phonology and syntax**: the relation between sound and structure. Cambridge: The MIT Press, 1984.

SENE, Marcus Garcia de; MASSINI-CAGLIARI, Gladis. Precisamos dar adeus ao achismo: o papel da Linguística no combate ao negacionismo. **Revista do Sell**, Uberaba, v. 12, n. 2, p. 90-113, 2023.

SERRA, Carolina Ribeiro. **Realização e percepção de fronteiras prosódicas no português do Brasil**: fala espontânea e leitura. 2009. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

SERRA, Carolina Ribeiro. A interface prosódia-sintaxe e o fraseamento prosódico no português do Brasil. **Journal of Speech Sciences**, Campinas, v. 5, n. 2, p. 47-86, 2016.

SILVA, Adelaide Hercília Pescatori. O estatuto da análise acústica nos estudos fônicos. **Cadernos de Letras da UFF – Dossiê: Letras e cognição**, Niterói, v. 41, p. 213-229, 2010.

SILVA, Adelaide Hercília Pescatori. Fazer fonética: ao IPA... e além! **Revista Versalete**, Curitiba, v. 8, n. 14, p. 320-344, jan./jun 2020a.

SILVA, Adelaide Hercília Pescatori. 42 é a resposta. Qual é a pergunta sobre a relação entre Computação e Linguística e tudo o mais? **SBC Horizontes**, Porto Alegre, ago. 2020b.

SILVA, Adelaide Hercília Pescatori; SILVA, Fabiano; CHACON, Lourenço. Scriba: uma inteligência artificial para auxiliar o ensino da ortografia. *In*: ALVES, Ubiratã Kickhöfel; MASSINI-CAGLIARI, Gladis (org.). **Fonologia e Ensino**: descobertas e interfaces. Campinas: Editora da Abralin, 2024. p. 75-104.

SILVERMAN, Kim; BECKMAN, Mary Esther; PITRELLI, John; OSTENDORF, Mori; WIGHTMAN, Colin; PRICE, Patti; PIERREHUMBERT, Janet Breckenridge; HIRSCHBERG, Julia. TOBI: a standard for labeling English prosody. *In*: 2ND INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 1992, Banff. **Proceedings** [...]. Banff, 1992. p. 867-870.

SILVESTRE, Aline Ponciano dos Santos. **A entoação regional dos enunciados assertivos nos falares das capitais brasileiras**. 2012. Tese (Mestrado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.

SILVESTRE, Aline Ponciano dos Santos. **"Se eu pudesse e se o meu dinheiro desse..."**: desgarramento e prosódia no português brasileiro e no português europeu. 2017. Tese (Doutorado em Letras Vernáculas) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017.

SILVESTRE, Aline Ponciano dos Santos; CUNHA, Cláudia de Souza. Pelos cantos do Brasil: a variação entonacional da asserção neutra em Natal, Rio de Janeiro e Porto Alegre. **Letrônica**, Porto Alegre, v. 6, n. 1, p. 179-195, jan./jun. 2013.

SIMÕES, Darcilia Marindir Pinto. **Iconicidade verbal**: teoria e prática. Rio de Janeiro: Dialogarts, 2009.

SONCIN, Geovana Carina Neris. Prosodic boundary in Brazilian Portuguese: the relation between auditory perception and phonetic cues. *In*: AGUIAR, Victoria Marrero; ESTEBAS VILAPLANA, Eva (org.). **Tendencias actuales en fonética experimental**. Madrid: UNED, 2017, p. 83-87.

SONCIN, Geovana Carina Neris; TENANI, Luciani Ester. Variações de F0 e configurações de frase entoacional: análise de estruturas contrastivas. **Domínios de Lingu@gem**, Uberlândia, v. 10, n. 2, p. 534-558, 2016.

SONCIN, Geovana Carina Neris; TENANI, Luciani Ester; BERTI, Larissa Cristina. Percepção de pausa em fronteira prosódica. **Scripta**, Belo Horizonte, v. 21, n. 41, p. 143-164, jun. 2017.

SONCIN, Geovana Carina Neris; TENANI, Luciani Ester; BERTI, Larissa Cristina. Phonologic representation and speech perception: the role of pause. **Diacrítica**: revista do Centro de Estudos Humanísticos, Braga, v. 33, n. 2, p. 4-18, 2019.

SOUZA, Carlos Francisco Soares de Souza. **Síntese de fala em português brasileiro baseada em modelos ocultos de Markov**. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2010.

STEFFMAN, Jeremy; JUN, Sun-Ah. Perceptual integration of pitch and duration: Prosodic and psychoacoustic influences in speech perception. **J. Acoust. Soc. Am.**, Melville, v. 146, n. 3, p. EL251-EL257, 2019.

TENANI, Luciani Ester. **Domínios prosódicos no português do Brasil**: implicações para a prosódia e para a aplicação de processos fonológicos. 2002. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2002.

TENANI, Luciani Ester. Fonologia Prosódica. *In*: HORA, Dermeval da; MATZENAUER, Carmen Lúcia (org.). **Fonologia, fonologias**: uma introdução. São Paulo: Contexto, 2017. p. 109-124.

TENANI, Luciani Ester; FERNANDES-SVARTMAN, Flaviane Romani. Prosodic phrasing and intonation in neutral and subject-narrow-focus sentences of Brazilian Portuguese. *In*: 4TH CONFERENCE ON SPEECH PROSODY, 2008, Campinas. **Proceedings** [...]. Campinas: RG/CNPq, 2008. p. 445-448.

TOJEIRA-RAMOS, Juan Prete. A caracterização prosódica da correlação consecutiva no português brasileiro. *In*: SILVA, Jacson Baldoino; DANTAS, Emanuelle Reisurreição; SILVA, Fagner Carvalho; ARAÚJO, Silvana Silva Farias de; SANTIAGO, Huda da Silva; BARREIROS, Liliane Lemos Santana; NASCIMENTO, Lucas (org.). **Diversidade linguística e práticas discursivas no contexto brasileiro**. Tutóia: Editora Lupa, 2024. p. 59-80.

TOJEIRA-RAMOS, Juan Prete. **Dicionário Pedagógico de Fonologia Entoacional do Português Brasileiro**. 2025. Dicionário digital. Lexonomy. Disponível em: <https://www.lexonomy.eu/35f8qhdy>. Acesso em: 10 de março de 2025.

TOJEIRA-RAMOS, Juan Prete; PEZATTI, Erotilde Goreti. A oração relativa apositiva no português escrito por adolescentes do noroeste paulista sob a abordagem discursivo-funcional. **Mosaico**, São José do Rio Preto, v. 20, n. 1, p. 259-282, 2021.

TOJEIRA-RAMOS, Juan Prete; PEZATTI, Erotilde Goreti. As propriedades prosódicas da oração relativa padrão sob a abordagem da Gramática Discursivo-Funcional. **Mosaico**, São José do Rio Preto, v. 21, n. 1, p. 143-171, 2022.

TOJEIRA-RAMOS, Juan Prete; GUIRALDELLI, Lisângela Aparecida. Análise prosódica das construções subordinadas predicativas finitas à luz da Gramática Discursivo-Funcional. **Mosaico**. São José do Rio Preto, v. 23, n. 1, p. 25-43, 2024.

TOJEIRA-RAMOS, Juan Prete; MASSINI-CAGLIARI, Gladis. Associação tonal no domínio prosódico da palavra fonológica no português brasileiro: o caso de dados produzidos por uma tecnologia de síntese de fala baseada em Inteligência Artificial. *In*: IX SIMPÓSIO MUNDIAL DE ESTUDOS EM LÍNGUA PORTUGUESA (SIMELP) e VI CONGRESSO DA ASSOCIAÇÃO INTERNACIONAL DE LINGUÍSTICA DO PORTUGUÊS (AILP), 2024, Funchal. **Anais** [...]. Funchal: Universidade da Madeira, [no prelo].

TONELI, Priscila Marques. **A palavra prosódica no Português Brasileiro**. 2014. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2014.

TONELI, Priscila Marques; VIGÁRIO, Marina; ABAURRE, Maria Bernadete Marques. Distinguishing emphatic and prosodic word initial stresses: evidence from Brazilian Portuguese. *In*: 4TH INTERNATIONAL SYMPOSIUM ON TONAL ASPECTS OF LANGUAGES (TAL), 2014, Nijmegen. **Proceedings** [...]. Nijmegen: ISCA, 2014. p. 172-176.

TONELI, Priscila Marques; ABAURRE, Maria Bernadete Marques; VIGÁRIO, Marina. Estrutura Entoacional de Sentenças Neutras em Português Brasileiro na variedade de Minas Gerais. **Filologia e Linguística Portuguesa**, São Paulo, v. 20, n. Especial, p. 47-70, 2018.

TRUCKENBRODT, Hubert; SANDALO, Maria Filomena Spatti; ABAURRE, Maria Bernadete Marques. Elements of Brazilian Portuguese intonation. **Journal of Portuguese Linguistics**, Lisboa, v. 8, p. 75-114, 2009.

TUNNERMANN, Daniel. **Controle de estilo na síntese de voz em português brasileiro usando redes neurais profundas**. 2021. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Goiás, Instituto de Informática, Goiânia, 2021.

VAN DEN OORD, Aäron; DIELEMAN, Sander; ZEN, Heiga; SIMONYAN, Karen; VINYALS, Oriol; GRAVES, Alex; KALCHBRENNER, Nal; SENIOR, Andrew; KAVUKCUOGLU, Koray. WaveNet: A Generative Model for Raw Audio. *In*: 9TH ISCA WORKSHOP ON SPEECH SYNTHESIS WORKSHOP, 2016, Sunnyvale. **Proceedings** [...]. Sunnyvale: ISCA, 2016. p. 125.

VIEIRA, Marcelo Augusto da Silva. **Acento tonal pré-nuclear ascendente no Português Brasileiro**: comparação com a fala disástrica parkinsoniana. 2017. Dissertação (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

VIGÁRIO, Marina. **The prosodic word in European Portuguese**. Berlin; New York: Mouton de Gruyter, 2003.

VIGÁRIO, Marina. O lugar do grupo clítico e da palavra prosódica composta na hierarquia prosódica: uma nova proposta. *In*: LOBO, Maria; COUTINHO, Maria Antónia (org.). **Actas do XXII Encontro Nacional da Associação Portuguesa de Lingüística**: textos seleccionados. Lisboa: Colibri Artes Gráficas, 2007. p. 673-688.

VIGÁRIO, Marina. Prosodic structure between the prosodic word and the phonological phrase: recursive nodes or an independent domain? **The Linguistic Review**, Berlin, v. 27, n. 4, p. 485-530, 2010.

VIGÁRIO, Marina; FERNANDES-SVARTMAN, Flaviane Romani. A atribuição de acentos tonais em compostos no português do Brasil. *In*: BRITO, Ana Maria; SILVA, Fátima; VELOSO, João; FIÉIS, Alexandra (org.). **XXV Encontro da Associação Portuguesa de Lingüística – Textos Seleccionados**. Porto: Tip. Nunes, 2010. v. 1, p. 769-786.

VON ZUBEN, Fernando J. Uma caricatura funcional de redes neurais artificiais. **Learning and Nonlinear Models – Revista da Sociedade Brasileira de Redes Neurais**, Rio de Janeiro, v. 1, n. 2, p. 66-76, 2003.