



Vinte anos do Projeto ALIP (Amostra Linguística do Interior Paulista) e o legado científico de Ataliba de Castilho

Twenty years of the Project ALIP (Linguistic Sample from São Paulo Countryside) and the legacy of Ataliba de Castilho

Cássio Florêncio RUBIO*

Sebastião Carlos Leite GONÇALVES**

RESUMO: O Banco de Dados Iboruna, desenvolvido no âmbito do Projeto ALIP (Amostra Linguística do Interior Paulista), completa 20 anos, destacando suas contribuições para a descrição do português falado no interior do estado de São Paulo. A trajetória desse Projeto dialoga diretamente com princípios da pesquisa linguística defendidos por Ataliba Teixeira de Castilho, especialmente o da valorização da língua em uso, da constituição de corpora representativos e da articulação entre descrição linguística e reflexão teórica. Com base nesse contexto, o objetivo desse artigo é realizar um balanço da produção científica desenvolvida a partir das amostras de fala do Banco de Dados Iboruna ao longo de suas duas décadas de existência, evidenciando sua amplitude temática e teórico-metodológica e suas contribuições para a descrição da variedade linguística do noroeste paulista e do Português Brasileiro. Com base em levantamento bibliográfico da produção científica que utiliza as amostras de fala do banco de dados como corpus, recortamos como referência resultados de pesquisas dedicadas a fenômenos fonético-fonológicos, morfossintáticos e textual-discursivos. O levantamento identificou mais de uma centena de trabalhos acadêmicos e, após a eliminação de sobreposições temáticas e objetos de estudo recorrentes, foram selecionadas 64 pesquisas representativas da produção desenvolvida com base no banco. Os estudos selecionados contemplam diferentes níveis de análise e evidenciam a amplitude temática e teórico-metodológica das investigações realizadas ao longo das duas décadas. Entre os trabalhos selecionados, predominam as pesquisas morfossintáticas, que correspondem a 62,5% do conjunto analisado, seguidas pelos estudos fonético-fonológicos, com 20,3%, e pelos estudos textual-discursivos, com 17,2%. A diversidade dos fenômenos investigados, que inclui processos de variação e mudança, demonstra a capacidade do Iboruna de sustentar diferentes agendas de investigação e de possibilitar análises da língua em uso a partir de uma base empírica comum. Os resultados evidenciam, ainda, a relevância do banco para a documentação da variedade linguística do noroeste paulista e do Português Brasileiro e para a formação de pesquisadores. Os vinte anos do Projeto ALIP e de seu banco de dados

* Professor Associado IV da Universidade Federal de São Carlos – UFSCar, Câmpus de São Carlos. cassiorubio@ufscar.br

** Professor Assistente Doutor da Universidade Estadual Paulista - Unesp, Câmpus de São José do Rio Preto. Bolsista de produtividade em Pesquisa do CNPq (Proc. 308248/2023-6). sebastiao.goncalves@unesp.br

representam não apenas a consolidação de um importante acervo linguístico, mas também a continuidade de um legado científico fundamental para a linguística brasileira. A produção desenvolvida com base no Iboruna demonstra sua capacidade de articular diferentes perspectivas teóricas e metodológicas e de contribuir para a descrição linguística fundamentada em dados de língua em uso.

PALAVRAS-CHAVE: Português Brasileiro. Português Paulista. Banco de Dados Iboruna. Ataliba Teixeira de Castilho.

ABSTRACT: The Iboruna Database, developed within the scope of the ALIP Project (Amostra Linguística do Interior Paulista / Linguistic Sample from the State of São Paulo Countryside), celebrates its 20th anniversary, highlighting its contributions to the description of Portuguese as spoken in the interior of the state of São Paulo. The trajectory of this project is closely aligned with principles of linguistic research advocated by Ataliba Teixeira de Castilho, particularly the valorization of language in use, the compilation of representative corpora, and the articulation between linguistic description and theoretical reflection. Against this background, this article aims to provide an overview of the scientific production developed from the speech samples in the Iboruna Database over its two decades of existence, highlighting its thematic and theoretical-methodological breadth and its contributions to the description of the linguistic variety spoken in northwestern São Paulo and of Brazilian Portuguese. Based on a bibliographic survey of scientific publications that use the database's speech samples as a corpus, we selected as a reference point findings from studies devoted to phonetic-phonological, morphosyntactic, and textual-discursive phenomena. The survey identified more than one hundred academic works and, after excluding thematic overlaps and recurring research topics, 64 studies representative of the research conducted using the database were selected. The selected studies address different levels of linguistic analysis and demonstrate the thematic and theoretical-methodological breadth of the research conducted over the two decades. Among the selected studies, morphosyntactic research predominates, accounting for 62.5% of the analyzed corpus, followed by phonetic-phonological studies, with 20.3%, and textual-discursive studies, with 17.2%. The diversity of the phenomena investigated, including processes of variation and change, demonstrates the Iboruna Database's capacity to support different research agendas and to enable analyses of language in use on the basis of a common empirical foundation. The results also highlight the database's relevance for documenting the linguistic variety of northwestern São Paulo and Brazilian Portuguese, as well as for researcher training. The twenty years of the ALIP Project and its database represent not only the consolidation of an important linguistic archive but also the continuation of a scientific legacy that is fundamental to Brazilian linguistics. The research developed using the Iboruna Database demonstrates its capacity to bring together different theoretical and methodological perspectives and to contribute to linguistic description grounded in data from language in use.

KEYWORDS: Brazilian Portuguese. Paulista Portuguese. Iboruna Database. Ataliba Teixeira de Castilho.

Artigo recebido em: 29.06.2026

Artigo aprovado em: 11.09.2026

1 Introdução

A constituição e o compartilhamento de bancos de dados linguísticos figuram entre as mais importantes transformações metodológicas ocorridas na linguística brasileira nas últimas décadas, principalmente por força do movimento *Ciência Aberta*, que, surgido na década de 1990 por meio do chamado *Open Access Movement*, tinha como principal objetivo provocar mudanças nos modelos vigentes de armazenamento, disseminação e visibilidade de pesquisas e dados científicos, bem como de acesso a eles (Albagli, 2015). O acesso a amostras representativas da língua em uso permitiu que o trabalho de descrição se tornasse cada vez mais fiel à realidade linguística, no que diz respeito ao conhecimento mais amplo do Português Brasileiro (PB), de suas variedades e dos fenômenos linguísticos de diferentes níveis de análise que os caracterizam.

Dentre os pesquisadores que mais contribuíram para essa mudança de paradigma no Brasil, destaca-se Ataliba Teixeira de Castilho, homenageado deste número temático. Ao longo de sua trajetória acadêmica, Castilho foi ferrenho defensor do trabalho coletivo e compartilhado de *corpora* linguísticos e de resultados de pesquisas sobre o PB, de forma a se alcançar uma visão mais ampla e real da língua tal como ela se manifesta socialmente (Castilho, 2021). Sua contribuição para a consolidação da pesquisa linguística no Brasil transcende suas reflexões teóricas individuais, materializando-se em sua liderança na articulação de grandes projetos coletivos nacionais. Castilho desempenhou um papel central na coordenação do *Projeto de Estudo da Norma Urbana Linguística Culta (NURC/SP)* e, subsequentemente, do *Projeto Gramática do Português Falado*, iniciativas pioneiras que revolucionaram a descrição da oralidade ao sistematizar a realidade do PB em uso (Castilho, 1990, 2021). Além disso, sua constante preocupação com os fundamentos históricos do PB culminou na idealização e estruturação do *Projeto Para a História do Português Brasileiro (PHPB)*, que impulsionou os estudos diacrônicos no país por meio da constituição de rigorosos *corpora* documentais (Castilho, 2018). Esses grandes empreendimentos

científicos não apenas pavimentaram o caminho para a compreensão empírica da nossa realidade linguística, mas também estabeleceram um modelo de cooperação acadêmica e formação de pesquisadores que define o cerne do legado de Castilho.

Foi por incentivo de Castilho que, no início dos anos 2000, nasce, na Universidade Estadual Paulista, Câmpus de São José do Rio Preto, o Projeto ALIP (Amostra Linguística do Interior Paulista) com seu banco de dados linguísticos denominado *Iboruna* (BDI, daqui em diante)¹. Concebido com o objetivo de documentar a variedade linguística falada na região noroeste do estado de São Paulo, nucleada por São José do Rio Preto e cidades circunvizinhas, o banco de dados se transformou em uma das mais importantes bases de dados para a investigação do PB do interior paulista, servindo de suporte metodológico para dezenas de dissertações, teses, artigos científicos, capítulos de livros e pesquisas de pós-doutorado. Trata-se de um banco de dados pioneiro no Brasil no acesso livre dado a toda sua documentação, do projeto que o concebeu até o material linguístico que o compõe (áudio das entrevistas, transcrições, diários de coleta e ficha social dos participantes), além de toda produção bibliográfica compilada e disponível no *site* do Projeto².

Embora possua características próprias e esteja vinculado a objetivos específicos de pesquisa, o BDI pode ser compreendido como parte de uma tradição científica mais ampla que encontra em Ataliba de Castilho um de seus principais formuladores e incentivadores (Castilho, 2017). A valorização da língua em uso, a preocupação com a representatividade dos dados, o interesse pela diversidade linguística brasileira e a defesa da articulação entre descrição e teoria constituem princípios que aproximam a trajetória do Projeto ALIP das ideias desenvolvidas pelo homenageado.

¹ O nome *Iboruna* ("Rio Preto", em tupi-guarani) foi imposto pelo Conselho Nacional de Geografia durante a reforma cartográfica da Era Vargas, para evitar homonímia com o município mineiro. Protestos locais levaram ao restabelecimento do nome original, *São José do Rio Preto*, sancionado pelo Decreto estadual 14.434 de 30/11/1944, que fixa a Divisão Administrativa e Jurídica do Estado de São Paulo (Gonçalves, 2005, 2007).

² Disponível em: <https://alip.ibilce.unesp.br/#> (Acesso em: 23 jun. 2026).

Ao completar vinte anos de existência, o BDI oferece uma oportunidade singular de reflexão sobre os avanços produzidos pela pesquisa linguística baseada em *corpora* no Brasil. Mais do que registrar a história de um acervo regional, trata-se de examinar os resultados de uma experiência científica que permitiu documentar aspectos centrais do português falado no interior paulista e formar sucessivas gerações de pesquisadores.

O presente artigo apresenta, portanto, um balanço da produção científica gerada a partir do BDI ao longo de suas duas décadas de existência. Com base em levantamento bibliográfico realizado em 2025, busca-se identificar as principais contribuições do BDI para a descrição do PB, bem como discutir seu significado no contexto da tradição de pesquisa linguística no Brasil, cuja consolidação contou decisivamente com a atuação de Castilho.

2 O Projeto ALIP, o Banco de Dados Iboruna e a tradição dos *corpora* linguísticos no Brasil

A história do Banco de Dados Iboruna não pode ser compreendida isoladamente da trajetória da Linguística brasileira nas últimas décadas. A partir da segunda metade do século XX, diferentes iniciativas voltadas à documentação sistemática da língua contribuíram para a formação de uma cultura de pesquisa baseada em *corpora*, permitindo que a descrição também do PB passasse a se apoiar cada vez mais em dados empiricamente observáveis.

Nesse cenário, a atuação de Ataliba de Castilho foi particularmente relevante. Seu envolvimento em projetos de documentação linguística (Castilho, 1990, 2018, 2021) e sua defesa constante da observação da língua em uso ajudaram a consolidar uma perspectiva segundo a qual a investigação linguística deve partir dos dados efetivamente produzidos pelos falantes. Para Castilho, compreender o PB significava observar sua história, sua diversidade, suas regularidades e seus processos de mudança em contextos reais de interação (Castilho, 2021).

O Projeto ALIP insere-se diretamente nessa tradição. Ao propor a constituição de um banco de dados com amostras linguísticas representativas da região de São José do Rio Preto, o projeto assumiu o compromisso de registrar empiricamente a fala do interior paulista, disponibilizando aos pesquisadores um conjunto de dados capaz de sustentar investigações em diferentes níveis de análise.

O desenho metodológico do BDI garante a sua robustez científica por meio de dois conjuntos amostrais. O primeiro, representativo do censo linguístico do interior paulista, reúne 151 entrevistas sociolinguísticas estratificadas por quatro variáveis censitárias - *sexo/gênero*, *escolaridade* (quatro níveis), *faixa etária* (cinco faixas) e *renda familiar* (quatro faixas) - e coletadas com base em roteiro de entrevista dirigido para obtenção de diferentes gêneros textual-discursivos (narrativo, argumentativo, descritivo, injuntivo). O segundo é composto de 11 interações dialógicas livres, coletadas de forma oculta, mas com a obtenção de consentimento posterior dos participantes. Essa configuração permite que os pesquisadores observem não apenas a ocorrência de determinados fenômenos linguísticos, mas também sua manifestação em contextos sociais reais, aspecto fundamental para a compreensão da língua em uso e dos processos de invariância, variância e mudança linguísticas.

A combinação entre estratificação social, diversidade de gêneros textual-discursivos e abrangência quantitativa das amostras tornou o BDI uma fonte privilegiada para caracterização do PB falado e, mais especificamente, da variedade do interior paulista, com pesquisas que abordam desde o nível fonético-fonológico ao textual-discursivo. Devido a essa versatilidade, o Projeto ALIP reuniu, ao longo de duas décadas, uma produção científica contínua que, em conjunto, podemos afirmar, consolida uma das mais detalhadas descrições já produzidas sobre uma variedade regional do português.

Outro aspecto que merece destaque se refere à diversidade teórico-metodológica de pesquisas para as quais o BDI vem se mostrando útil como recurso metodológico. Servindo não somente a pesquisas do âmbito da Sociolinguística

variacionista, o BDI subsidia, com igual eficácia, investigações vinculadas a diferentes correntes da Linguística contemporânea, incluindo o Funcionalismo, a Linguística Textual, a Pragmática, a Gramática das Construções e suas abordagens cognitivo-funcionalistas, servindo, ainda, mais recentemente, a outras áreas de interface com a Linguística, como a que investiga o Processamento de Linguagem Natural para o Português³. Essa pluralidade permite que um mesmo conjunto de dados seja explorado sob múltiplas perspectivas analíticas, ampliando significativamente o alcance científico do Projeto ALIP e de seu banco de dados.

Assim, ao completar vinte anos de existência, o BDI apresenta um legado que ultrapassa a própria documentação da fala regional. Sua principal contribuição reside na articulação entre documentação linguística, formação de pesquisadores e produção continuada de conhecimento científico, permitindo que diferentes gerações de estudiosos construam, de forma cumulativa, um retrato multifacetado do português falado no interior paulista.

3 Metodologia

Este artigo é um recorte de um projeto mais amplo, em andamento, voltado à elaboração de uma obra de caráter historiográfico e científico destinada a reunir, organizar e divulgar os resultados alcançados ao longo de duas décadas de pesquisas desenvolvidas a partir do BDI. Inicialmente foram identificados mais de uma centena de trabalhos acadêmicos que utilizaram as amostras de fala disponibilizadas no banco de dados como recurso metodológico. Após a eliminação de sobreposições temáticas e de repetição de objetos de estudo, selecionamos 64 pesquisas representativas do trabalho de descrição da fala do interior paulista.

³ Amostras de fala do BDI vêm servindo ao grupo de pesquisa de Processamento de Língua Natural para o Português (NLP2), abrigado no C4AI - *Center of Artificial Intelligence*, sediado na Universidade de São Paulo. (cf. <https://c4ai.inova.usp.br/>. Acesso em: 23 jun. 2026).

Os trabalhos foram organizados de acordo com os níveis linguísticos investigados, permitindo a identificação de três grandes eixos de análise: estudos fonético-fonológicos, estudos morfossintáticos e estudos textual-discursivos. Em cada caso, foram considerados o objeto de pesquisa, os referenciais teóricos adotados, os procedimentos metodológicos empregados e os principais resultados alcançados.

O Quadro 1 apresenta os principais fenômenos investigados em cada um dos níveis de análise identificados no levantamento bibliográfico.

Quadro 1 – Principais fenômenos investigados por nível de análise, utilizando o BDI.

Nível de análise	Fenômenos investigados
Fonético-fonológico	Haplologia, Alçamento vocálico, Síncope vocálica, Harmonia vocálica, Redução de gerúndio, Apagamento vocálico, Metátese, Sândi vocálico, Rotacismo.
Morfossintático	Concordância verbal, Concordância nominal, Expressão do sujeito, Alternância pronominal, Gramaticalização e variação de Tempo, Modo e Aspecto, Formas de tratamento, Combinação de orações, Modalidade e Evidencialidade.
Textual-discursivo	Marcadores discursivos, Referenciação, Hesitação, Organização narrativa, Construções subjetivas, Expressividade e avaliação, Organização tópica.

Fonte: elaborado pelos autores.

A Tabela 1 apresenta a distribuição quantitativa dos 64 trabalhos em análise. Os dados revelam o predomínio das pesquisas de natureza morfossintática, que correspondem a 62,5% da produção identificada. Em seguida, aparecem os estudos fonético-fonológicos, com 20,3%, e os de natureza textual-discursiva, responsáveis por 17,2% do total.

Tabela 1 – Quantitativo de pesquisas por nível de análise.

Nível de análise	Número de trabalhos	Percentual
Fonético-fonológico	13	20,3%
Morfossintático	40	62,5%
Textual-discursivo	11	17,2%
Total	64	100%

Fonte: elaborada pelos autores.

Em conjunto, os resultados demonstram a amplitude das investigações desenvolvidas com base no BDI e confirmam sua relevância para a descrição do PB falado e para a consolidação de diferentes agendas de pesquisa linguística.

Nossa opção por destacar, neste artigo, alguns trabalhos mais representativos respeita os limites impostos pelo próprio gênero. Buscamos construir uma síntese interpretativa que seja capaz de evidenciar a relevância do conjunto das pesquisas desenvolvidas ao longo das duas décadas de existência do banco.

4 O retrato linguístico do interior paulista

4.1 Estudos fonético-fonológicos

Os estudos fonético-fonológicos desenvolvidos a partir do BDI constituem uma das vertentes mais tradicionais das pesquisas vinculadas ao Projeto ALIP. Ao longo de duas décadas, diferentes investigações dedicaram-se à descrição de processos variáveis que afetam a estrutura sonora do português falado no noroeste paulista, contemplando fenômenos relacionados ao sistema vocálico, à redução segmental, à reorganização silábica e aos aspectos prosódicos da fala. Em conjunto, essas pesquisas permitiram documentar características da variedade regional e contribuir para debates mais amplos sobre variação e mudança linguística no PB.

Uma das descrições mais produtivas abarca o comportamento das vogais médias átonas. Nesse âmbito, Silveira (2008) analisou o alçamento das vogais médias pretônicas em substantivos, demonstrando que a harmonização vocálica desempenha papel central na elevação das vogais /e/ e /o/ em palavras como *aligria*, *piqueno* e *nutícia*. A autora verificou que a presença de vogais altas na sílaba subsequente favorece

significativamente a aplicação da regra variável, aproximando os resultados obtidos daqueles observados em outras regiões do país. Já Carmo (2009) voltou-se para a investigação do mesmo fenômeno em formas verbais, examinando ocorrências como *prifiro*, *cunheço* e *cumeçar*. Seus resultados demonstraram que, embora o alçamento apresente condicionamentos fonológicos semelhantes aos encontrados nos substantivos, fatores morfológicos próprios da classe verbal, como conjugação verbal, também exercem influência sobre a distribuição das variantes. Posteriormente, Carmo (2013) ampliou sua investigação anterior, ao considerar conjuntamente nomes e verbos, confirmando a harmonização vocálica como o principal mecanismo responsável pelo alçamento das vogais pretônicas na variedade estudada.

Ainda no âmbito das vogais átonas, Ramos (2009) examinou o comportamento das vogais médias postônicas não finais. Além de documentar ocorrências de alçamento em formas como *númiro*, *tíquite* e *épuca*, a autora investigou também a síncope vocálica em palavras proparoxítonas. O estudo registrou realizações como *numro* (número), *abóbra* (abóbora), *árvre* (árvore), *máscra* (máscara), *dizmo* (dízimo) e *pêsgo* (pêssego), demonstrando que o apagamento da vogal postônica não final constitui um fenômeno relativamente raro na comunidade analisada, com apenas 8% de aplicação da regra. Os resultados indicaram que o processo é favorecido quando a vogal é seguida por líquidas, nasais ou sibilantes, evidenciando a relevância da estrutura silábica para sua implementação.

Outro conjunto importante de pesquisas dedicou-se aos processos de redução segmental. Nessa direção, Ferreira (2010) investigou o apagamento do fonema /d/ em morfemas de gerúndio, fenômeno amplamente difundido no PB falado. A autora analisou quase mil ocorrências extraídas de amostras de fala do BDI e verificou que a forma reduzida é predominante, alcançando 72% dos casos observados. Formas como *andano*, *chorano*, *voano* e *pulano* mostraram-se particularmente frequentes, sobretudo entre homens, falantes mais jovens e indivíduos com menor escolaridade. O estudo demonstrou ainda que o apagamento está associado a alterações articulatorias que

reduzem a perceptibilidade do segmento consonantal /d/, configurando um processo simultaneamente fonológico e sociolinguístico.

Outros tipos de reorganização da estrutura silábica também despertaram interesse dos pesquisadores. Pavezzi (2006) analisou a haplologia, processo pelo qual uma sílaba é eliminada em sequências foneticamente semelhantes, registrando ocorrências como *denda viatura* (por *dentro da viatura*). Os resultados evidenciaram que a aplicação da regra depende de fatores fonológicos específicos e de condicionamentos prosódicos, contrariando interpretações que atribuíam ao fenômeno um caráter meramente ocasional. Em linha semelhante, Dias e Carmo (2021) investigaram a metátese, observando formas como *dromir* e *sastifação*. Os autores demonstraram que a troca de posição entre segmentos ocorre de maneira sistemática e está fortemente associada à presença do rótico /r/, responsável pela grande maioria dos casos identificados.

Uma contribuição particularmente inovadora para os estudos fonético-fonológicos desenvolvidos com dados do BDI encontra-se na pesquisa de Marcato (2013), dedicada ao comportamento prosódico das preposições monossilábicas e da preposição *para*. Fundamentada na Fonologia Prosódica, a autora analisou mais de 12 mil ocorrências e demonstrou que esses elementos se comportam como clíticos, prosodizando-se preferencialmente com a palavra subsequente. A autora identificou diversos processos fonológicos em atuação, dentre eles: a assimilação nasal (*com muito* > *ku muito*), neutralização vocálica (*de* > *di*), redução fonológica (*para* > *pra* > *pa*), metátese (*em nós* > *ni nós*) e diferentes tipos de sândi vocálico externo. Os resultados revelaram que mais de 66% dos contextos potencialmente favoráveis apresentavam algum tipo de processo de juntura, evidenciando a importância dos fatores prosódicos na organização da fala espontânea.

Em conjunto, os estudos fonético-fonológicos produzidos a partir do BDI revelam uma agenda de pesquisa diversificada e teoricamente plural, que articula pressupostos da Sociolinguística Variacionista, da Fonologia Métrica, da Fonologia

Prosódica e de diferentes modelos fonológicos contemporâneos. Mais do que descrever fenômenos específicos da fala do interior paulista, essas investigações contribuíram para o entendimento dos mecanismos de variação e mudança linguística do PB, reafirmando a importância dos *corpora* de língua falada para a construção de descrições empiricamente fundamentadas da língua em uso.

4.2 Estudos morfossintáticos

Os estudos morfossintáticos constituem o eixo mais representativo da produção científica desenvolvida no interior do Projeto ALIP. Como apontado previamente, dos 64 trabalhos identificados no levantamento, 40 (62,5%) concentram-se nesse nível de análise, evidenciando a centralidade das questões gramaticais na trajetória do Projeto ALIP.

Ao longo de duas décadas, as pesquisas contemplaram fenômenos relacionados à concordância verbal e nominal, à expressão do sujeito, à alternância pronominal, aos modos e tempos verbais, às formas de tratamento, à modalidade, à evidencialidade e aos processos de gramaticalização. Em conjunto, esses estudos revelam a forte influência da Sociolinguística Variacionista na constituição do banco de dados, ao mesmo tempo em que demonstram a incorporação progressiva de perspectivas funcionalistas e cognitivo-discursivas.

Entre os temas mais recorrentes se encontra a concordância verbal. Um dos estudos pioneiros desenvolvidos com dados do Iboruna foi realizado por Rubio (2008), que investigou a concordância verbal de terceira pessoa do plural, em ocorrências como *Só pá quem as vaca **conhece** assim* e *Alguns dos mais antigos **contavam** por exemplo DO... PAdre*. A pesquisa demonstrou que a presença da marca de plural no verbo constitui uma regra variável, influenciada por fatores linguísticos e sociais. Entre os condicionadores mais relevantes se destacaram a saliência fônica da oposição singular/plural, a posição do sujeito em relação ao verbo, o paralelismo formal e o nível de escolaridade dos falantes. O autor confirma que a variedade analisada apresenta

comportamento semelhante ao observado em outras comunidades de fala brasileiras, reafirmando a natureza variável da concordância verbal no português falado.

Essa linha de investigação foi posteriormente ampliada por Rubio (2012), em estudo comparativo entre o PB e o Português Europeu. Ao relacionar concordância verbal e alternância pronominal, o autor demonstrou que as transformações observadas no quadro pronominal repercutem diretamente sobre os padrões flexionais do verbo. A análise contemplou, além da terceira pessoa do plural, construções de primeira pessoa envolvendo formas como *nós* e *a gente*, como segue: ***nós trabalha** numa base de umas cento e cinquenta pessoa; nós **somo(s)** um país muito Rico em petróleo; a gente se **conheceu** há uns dois anos e nesse ano a gente **fomo(s)** pra lá né?*. A análise evidenciou que a expansão de pronomes inovadores favorece a reorganização dos paradigmas de concordância e que mudanças tradicionalmente analisadas de forma isolada, como a substituição de *nós* por *a gente* ou a expansão de *você*, integram um processo mais amplo de reestruturação morfossintática do PB contemporâneo.

Ainda no domínio da concordância, Fiamengui (2011) examinou a concordância nominal na fala e na escrita de adolescentes da região de São José do Rio Preto. A autora demonstrou que a marcação de plural em sintagmas nominais também constitui um fenômeno variável condicionado por fatores estruturais e sociais. A análise de ocorrências como *ai chegou os **parente** dele* e *ai tinha **alguns vidro quebrado*** evidenciou que a distribuição das marcas flexionais depende da configuração interna do sintagma, da modalidade de uso da língua e do perfil sociocultural dos falantes. Os resultados aproximam-se daqueles encontrados em outras variedades do PB e reforçam a ideia de que a concordância nominal integra um conjunto amplo de processos variáveis presentes na gramática da língua.

A concordância em estruturas predicativas foi objeto da pesquisa de Salomão (2010), que investigou construções envolvendo predicativos nominais e adjetivais, como nas ocorrências que seguem: *na época éh... as coisa era muito **rígida*** e *os alunos já estão **alfabetizados***. O estudo revelou que a concordância nesses contextos também

está sujeita à variação, sendo favorecida por fatores como a presença de marcas anteriores de plural, a posição dos constituintes e a saliência das marcas flexionais. Ao demonstrar que mesmo estruturas frequentemente tratadas como categóricas pela tradição gramatical apresentam comportamento variável na fala espontânea, a autora ampliou o alcance dos estudos de concordância desenvolvidos a partir do BDI.

Os modos verbais também receberam atenção expressiva. Entre os estudos pioneiros, destaca-se a pesquisa de Santos (2005) sobre a alternância entre indicativo e subjuntivo em orações subordinadas (*se eu fosse a professora... eu **acharia** um método mais fácil de não gritar e eu achava que eu **queria** fisioterapia*). A autora demonstrou que a escolha entre os dois modos não depende exclusivamente das prescrições normativas, sendo influenciada por fatores semânticos, pragmáticos e discursivos. Os resultados indicaram que o subjuntivo tende a ser favorecido em contextos associados à dúvida, à hipótese e à volição, enquanto o indicativo ocorre com maior frequência em ambientes marcados pela factualidade e pela certeza. O estudo contribuiu para uma compreensão mais refinada da modalidade verbal e evidenciou a importância dos fatores de uso na seleção modal.

A investigação dos tempos verbais foi ampliada por Brandão (2011), que analisou a alternância entre o pretérito imperfeito e o futuro do pretérito em construções condicionais. Ao examinar ocorrências como *se tivesse pego ele **poderia** num tá entre nós e se ela num tivesse namoran::(d)o assim a gente **podia** í(r) no cine::ma*, a autora verificou que as duas formas coexistem na fala contemporânea, sendo condicionadas por fatores semânticos e discursivos. Os resultados revelaram que a escolha entre as variantes está relacionada ao grau de hipoteticidade, ao contexto comunicativo e ao envolvimento do falante com o conteúdo enunciado, evidenciando a natureza dinâmica do sistema temporal do português.

Ainda no domínio verbal, Fonseca (2010) investigou a construção perifrástica *ir + infinitivo* na expressão do futuro, como em *um dia eu **vô(u) morrê(r)**... e **vô(u) tê(r)** a eternidade*. Seus resultados mostraram, para a variedade falada no interior paulista, a

total obsolescência do futuro sintético, restrito a expressões formulaicas (*será que ...*) e a certos aforismos bíblicos, reforçando observações já registradas em outras variedades do PB. A pesquisa apontou ainda a influência de fatores discursivos, como função modal, e pragmáticos, como função de marcador discursivo, na escolha entre as formas concorrentes.

Outro conjunto significativo de pesquisas voltou-se para a análise de estruturas complexas e processos de encaixamento oracional. Nessa linha, Santana (2010) investigou construções subordinadas sob uma perspectiva funcionalista, demonstrando que as relações sintáticas estabelecidas entre as orações refletem necessidades comunicativas específicas dos falantes. Posteriormente, Andrade (2016) examinou construções de alçamento de sujeito e de objeto à posição sujeito, como em *eu sou difícil de aprender* e em *toalha é complicado dobrar*, mostrando tais construções desempenham importantes funções discursivas e cognitivas, nas estratégias de organização da informação.

A modalidade e a evidencialidade constituem temas de investigação importantes da descrição morfossintática derivada do BDI. Vendrame (2010) investigou os verbos de percepção *ver*, *ouvir* e *sentir* (como em *eu vi uma... sabe? senti uma MÃO... puxan(d)o com TU-DO* e *Senti que a briga estava mais violenta*), demonstrando que essas formas podem ultrapassar seus significados lexicais originais para desempenhar funções relacionadas à expressão da fonte da informação e ao grau de comprometimento do falante com o conteúdo enunciado. Em perspectiva semelhante, Kapp-Barboza (2017) analisou os usos do verbo *saber* e constatou que essa forma verbal participa de diferentes estratégias de marcação epistêmica e evidencial, como nas ocorrências *Faz tempo que eu num faço eu sei que::... eu refogava* e *soube pelo barulho da conversa e o cheiro bom da comida da minha mãe*. Os resultados revelaram a complexidade das relações entre gramática, cognição e interação, aproximando as pesquisas desenvolvidas no âmbito do Iboruna de discussões contemporâneas sobre subjetividade e construção de sentidos.

Dentre os estudos de orientação funcionalista desenvolvidos com amostras de fala do BDI, destaca-se a pesquisa de Fortilli (2013) sobre os predicados matrizes adjetivais de orações subjetivas. A autora investigou processos de gramaticalização e dessentencialização de oração envolvendo construções tradicionalmente formadas por cópula, adjetivo e oração completiva, demonstrando que determinadas estruturas vêm passando por processos de reanálise sintático-semântico na língua falada. Os resultados revelaram a coexistência de construções mais conservadoras, como *é bom eu prestar uma universidade pública* e *é lógico que ele percebeu*, com formas inovadoras, sem cópula, como em *lógico que é tudo moderno*, e com função adverbial, como em *é tudo moderno... lógico*. Conforme os resultados da autora, adjetivos de natureza epistêmica e avaliativa tendem a perder gradativamente características sintáticas da estrutura original e a assumir funções cada vez mais relacionadas à expressão da atitude do falante diante da proposição, evidenciando trajetórias de mudança que aproximam gramática e discurso.

Mais recentemente, Almeida (2020) investigou os usos não locativos do item *onde*, como em *tipo assim... aí era onde eu já num aceitava e falava pra ela*. Tradicionalmente descrito como advérbio relativo de lugar, os usos de *onde* revelaram ampla expansão funcional na fala contemporânea, ocorrendo em contextos que extrapolam a referência espacial concreta. Os resultados mostraram que construções como essas refletem processos de mudança construcional em andamento, pelos quais determinados itens linguísticos ampliam seu escopo de atuação e passam a integrar novos domínios semânticos e discursivos.

Tomados em conjunto, os estudos morfosintáticos desenvolvidos a partir do BDI revelam uma agenda de pesquisa ampla e diversificada, marcada pela convergência entre descrição empírica, análise quantitativa e reflexão teórica. As investigações sobre concordância, sistema pronominal, modalidade, tempos verbais, encaixamento oracional e gramaticalização permitiram documentar importantes

aspectos da gramática do português contemporâneo, evidenciando processos de variação e mudança em pleno andamento.

4.3 Estudos textual-discursivos

Embora representem uma parcela menor da produção científica derivada do BDI, os estudos textual-discursivos desempenham papel fundamental na ampliação das perspectivas analíticas mobilizadas pelas pesquisas vinculadas ao Projeto ALIP. Ao longo de duas décadas, as investigações desenvolvidas nesse eixo passaram a explorar diferentes aspectos da organização do discurso falado, contemplando temas como marcadores discursivos, estruturação tópica, tradições discursivas, subjetividade, metadiscursividade, expressividade e hesitação. Em conjunto, esses trabalhos contribuíram para demonstrar que a fala espontânea constitui um objeto complexo de investigação, marcado por mecanismos sofisticados de organização textual e interação verbal.

Entre os estudos pioneiros destaca-se a pesquisa de Guerra (2007) sobre as funções textual-interativas de marcadores discursivos (MD). Ao analisar MD como *né*, *então*, *aí*, *daí* e *sabe*, em construções como *aí no outro dia...* *aí minha filha falou...* e *o que vale hoje é o dinheiro né?*, a autora descreveu funções essenciais dessa categoria na construção do texto falado. Mais do que simples conectores ou recursos de preenchimento, os MD atuam na introdução e manutenção de tópicos, na organização sequencial dos enunciados, no gerenciamento dos turnos conversacionais e na negociação de sentidos entre os interlocutores. Os resultados evidenciaram a centralidade desses elementos na fluidez do discurso e na coerência da interação verbal.

A organização tópica do discurso foi aprofundada por Penhavel (2010), que investigou os mecanismos responsáveis pela progressão temática em relatos de opinião das entrevistas sociolinguísticas do BDI. O autor mostra que a construção de tópicos discursivos depende da articulação entre diferentes recursos linguísticos, entre

eles os MD, as estruturas sintáticas e as estratégias argumentativas. A pesquisa revelou como os falantes introduzem, desenvolvem, retomam e encerram tópicos ao longo da interação verbal, contribuindo para uma compreensão mais detalhada da arquitetura do texto oral.

Outra importante pesquisa é a de Lopes-Damásio (2011), que examinou a relação entre tradições discursivas e mudança linguística. Partindo da observação de usos recorrentes presentes nas entrevistas do Iboruna, a autora constatou que determinados padrões discursivos favorecem a estabilização e a difusão de novas funções linguísticas. Seus resultados evidenciaram a relevância das práticas discursivas para a compreensão dos processos de mudança e para a constituição de novos usos na língua, aproximando as pesquisas vinculadas ao Projeto ALIP dos debates contemporâneos sobre historicidade textual e gramática emergente.

Também merece destaque a pesquisa de Caixeta (2015) sobre as interjeições no português falado. Ao examinar ocorrências como em *vixe aquele monte de gente* e em *NOSSA... chega nas faculdade eles aparece mesmo*, o autor demonstrou que as interjeições não constituem elementos periféricos do discurso, mas recursos importantes para a expressão de emoções, avaliações e reações dos falantes. O estudo evidenciou como as interjeições participam da construção da expressividade e da organização retórica dos textos orais.

Fortilli (2016) examinou os enunciados metadiscursivos presentes em entrevistas do BDI focalizando os mecanismos pelos quais os falantes organizam e monitoram o próprio discurso durante a interação verbal. Com base na perspectiva textual-interativa, a autora demonstrou que expressões como as em destaque em *Ah, eu queria falar um pouco sobre um tema...*, *Bom... eu vou falar de um evento que aconteceu na minha vida...*, *Mas deixa eu terminar a história...* e *Bom, então eu posso falar de outro aluno agora?* desempenham papel fundamental na introdução de tópicos, na reformulação de informações, na negociação dos turnos conversacionais e na condução da interação entre entrevistador e entrevistado. Os resultados evidenciaram que a entrevista

sociolinguística constitui um espaço de intensa atividade reflexiva, no qual os participantes utilizam estratégias metadiscursivas para planejar a fala, garantir a progressão textual e adequar continuamente sua produção às exigências comunicativas da situação de interação.

Nagamura (2016) analisou recursos linguísticos empregados na manifestação de avaliações subjetivas e posicionamentos discursivos. Os resultados mostraram que a subjetividade permeia diferentes níveis da organização textual, influenciando a seleção de formas linguísticas, a construção argumentativa e a forma como os falantes negociam perspectivas e pontos de vista ao longo da interação verbal.

Outra contribuição importante se encontra no estudo de Wiedemer (2019) sobre hesitação na fala. A análise de pausas preenchidas, repetições, alongamentos e reformulações, como a que se verifica na ocorrência *então éh::... ela é uma pessoa assim muito amor::sa*, revelou que esses fenômenos não representam simples falhas de produção, mas mecanismos que auxiliam o planejamento do discurso, a manutenção do turno conversacional e a organização cognitiva da fala. Os resultados contribuíram para uma compreensão mais refinada da oralidade e dos processos envolvidos na construção textual em tempo real.

Considerados em conjunto, os estudos textual-discursivos com base nas amostras de fala do BDI evidenciam a ampliação progressiva das agendas de pesquisa vinculadas ao Projeto ALIP. Dos primeiros trabalhos sobre marcadores discursivos e organização tópica às investigações mais recentes sobre subjetividade, metadiscursividade e hesitação, observa-se um crescente interesse pelos mecanismos responsáveis pela construção dos sentidos e pela organização do texto falado construído na interação verbal. Essa produção demonstra a versatilidade do banco de dados e sua relevância para diferentes correntes dos estudos do texto e do discurso, reafirmando a importância da língua falada como objeto privilegiado de investigação linguística.

5 Contribuições científicas do Banco de Dados Iboruna: um projeto em sintonia com o legado de Ataliba de Castilho

Ao trazer à mostra o conjunto de pesquisas desenvolvidas com base no BDI torna-se evidente a convergência entre os objetivos do Projeto ALIP, que o concebeu, e princípios que marcam a trajetória intelectual de Ataliba de Castilho.

Um desses princípios refere-se à centralidade do dado empírico na construção do conhecimento linguístico. Ao longo de sua atuação na Linguística, Castilho insiste na necessidade de fundamentar as análises em amostras representativas da língua em uso. Os estudos que recorrem ao BDI materializam essa perspectiva ao investigar fenômenos linguísticos concretos observados em situações reais de comunicação.

Outro aspecto de aproximação encontra-se na valorização da diversidade linguística brasileira. Assim como Castilho sempre defendeu a importância de compreender a pluralidade constitutiva do PB falado no Brasil, o BDI tornou possível a descrição detalhada de uma variedade regional específica, contribuindo para ampliar o conhecimento sobre a heterogeneidade linguística do país.

Mais do que a soma dos trabalhos individualmente considerados, o levantamento realizado evidencia a extraordinária capacidade do BDI de gerar conhecimento científico de forma contínua ao longo de mais de vinte anos. Os 64 estudos identificados, abordando os níveis fonético-fonológico, morfossintático e textual-discursivo, demonstram não apenas a diversidade temática das pesquisas desenvolvidas, mas também a consolidação de uma agenda permanente de investigação sobre o PB falado no interior paulista.

Essa continuidade permitiu a construção progressiva de um amplo retrato da variedade linguística do noroeste paulista. Considerados em conjunto, os estudos assentados metodologicamente no BDI oferecem descrições detalhadas de processos variáveis, mudanças linguísticas em curso e estratégias de organização da interação verbal. O resultado é um conjunto de conhecimentos que ultrapassa os objetivos específicos de cada investigação e contribui para a caracterização abrangente de uma das variedades mais bem descritas do PB contemporâneo.

Além disso, o levantamento evidencia que a vitalidade científica do BDI permanece ativa. Parte significativa das pesquisas identificadas foi desenvolvida durante a segunda década de existência do Projeto ALIP, contemplando temas relacionados à subjetividade, metadiscursividade, construcionalização, gramaticalização, modalidade e cognição. Esse movimento demonstra que o Iboruna não apenas preserva sua relevância como fonte de dados, mas também continua alimentando investigações alinhadas a agendas contemporâneas da Linguística Funcional, da Linguística Cognitiva e dos estudos do discurso.

Sob essa perspectiva, o principal legado do BDI talvez resida justamente em sua capacidade de articular documentação linguística, formação de pesquisadores e produção continuada de conhecimento científico. Ao longo de seus vinte anos, o BDI serviu de base para dezenas de trabalhos de iniciação científica, dissertações, teses e publicações, contribuindo simultaneamente para o avanço dos estudos sobre o PB e para a formação de novas gerações de pesquisadores. Trata-se, portanto, não apenas de um banco de dados, mas de um projeto científico de longa duração que continua produzindo resultados relevantes para a Linguística brasileira. Sua principal contribuição não reside apenas na documentação linguística de uma comunidade de fala específica, mas também na criação de condições para que diferentes tradições teóricas dialoguem a partir de uma mesma base empírica. Poucos *corpora* brasileiros produziram um conjunto tão diversificado de pesquisas, capazes de iluminar diferentes dimensões da estrutura e do funcionamento da língua. O resultado é um retrato multifacetado do português do interior paulista, construído coletivamente ao longo de pouco mais de vinte anos de investigação científica contínua.

6 Considerações finais

Ao completar vinte anos de existência, o BDI consolida-se como um dos mais robustos *corpora* de língua falada produzidos no Brasil. As dezenas de pesquisas desenvolvidas a partir de seus dados demonstram sua relevância para a descrição do

português falado no interior paulista, para a formação de pesquisadores e para o avanço do conhecimento linguístico.

Mais do que um banco de dados regional, o BDI representa a concretização de uma concepção de pesquisa que marcou profundamente a linguística brasileira contemporânea. Sua trajetória reafirma a importância da documentação sistemática da língua, da valorização dos dados empíricos e do reconhecimento da diversidade linguística como objeto legítimo de investigação científica. Por essa razão, celebrar seus vinte anos significa também reconhecer a permanência e a atualidade do legado intelectual de Ataliba Teixeira de Castilho. Em diferentes momentos de sua trajetória, o linguista defendeu a construção de *corpora* representativos, a observação rigorosa da língua em uso e o diálogo permanente entre teoria e descrição. O percurso do Iboruna demonstra a vitalidade dessa agenda científica e evidencia como seus princípios continuam produzindo resultados relevantes para os estudos linguísticos.

Assim, ao mesmo tempo em que registra duas décadas de contribuições para os estudos sobre o PB, este artigo presta homenagem a um pesquisador cuja atuação ajudou a criar as condições intelectuais e metodológicas que tornaram possíveis iniciativas como a do Projeto ALIP com seu banco de dados, cuja história, nesse sentido, é também parte da história da própria linguística brasileira contemporânea.

Referências

ALBAGLI, S. Ciência aberta em questão. In: ALBAGLI, S.; MACIEL, M. L.; HANNUD, A. (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. p. 9-26. DOI <https://doi.org/10.18225/978-85-7013-109-6>

ALMEIDA, M. A. “Onde está você agora?": uma investigação sobre os contextos não locativos de *onde* na fala paulista. **Revista Falange Miúda**. v. 5, n. 2, 2020, p. 157-175. Disponível em: <https://periodicos.upe.br/index.php/refami/article/view/427>. Acesso em: 18 jun. 2025.

ANDRADE, G. S. **Investigação funcionalista do alçamento de constituintes argumentais à posição de sujeito**. 2016. 111f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2016. Disponível

em: <https://repositorio.unesp.br/entities/publication/13556661-5a99-41b5-b111-5af9cd1beb43>. Acesso em: 22 abr. 2025.

BRANDÃO, S. M. **Variação em formas verbais**: um estudo sociolinguístico da alternância entre futuro do pretérito e pretérito imperfeito do indicativo no português paulista. 2015. 87f. Trabalho de conclusão de curso (Bacharelado em Letras) – Universidade Estadual Paulista, Faculdade de Ciências e Letras, Araraquara, 2015. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/b8fdc05f-8325-438d-a8e2-03ed4a5a3899/content>. Acesso em: 15 abr. 2025.

BRANDÃO, S. M. **Alternância verbal em construções condicionais**: um fenômeno variável?. 2018. 146f. Dissertação (Mestrado em Linguística e Língua Portuguesa) – Universidade Estadual Paulista, Faculdade de Ciências e Letras, Araraquara, 2018. Disponível em: <https://repositorio.unesp.br/entities/publication/36fb27db-4f40-40ee-9fc7-a45f761545dc>. Acesso em: 15 abr. 2025.

CAIXETA, G. F. **Que bom, que bom, ai, que bom!** Da existência da relação retórica de interjeição. 2015. 261f. Tese (Doutorado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Faculdade de Letras, Belo Horizonte, 2015. Disponível em: <https://repositorio.ufmg.br/handle/1843/MGSS-A72IP5>. Acesso em: 22 mai. 2025.

CARMO, M. C. **As vogais médias pretônicas dos verbos na fala culta do interior paulista**. 2009. 119 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2009. Disponível em: https://repositorio.unesp.br/entities/publication/297571b9-7496-4377-8907-dfda710db555?utm_source=chatgpt.com. Acesso em: 22 abr. 2025.

CARMO, M. C. **As vogais médias pretônicas na variedade do interior paulista**. 2013. 248 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2013. Disponível em: <https://repositorio.unesp.br/entities/publication/d3d4abbb-0b7a-4e77-ad9f-005146c780ff>. Acesso em: 22 abr. 2025.

CASTILHO, A. T. Gramática do português falado: fundamentos e perspectiva. **Cadernos de Linguística**, Campinas, n. 1, v. 2, p. 01-17, 2021. DOI <https://doi.org/10.25189/2675-4916.2021.v2.n1.id252>

CASTILHO, A. T. (coord.). **História do Português Brasileiro**: o português em seu contexto histórico. São Paulo: Editora Contexto, 2018.

CASTILHO, A. T. A UNESP e a linguística brasileira. **Estudos Linguísticos**, São Paulo, n. 46, v. 1, p. 109-137, 2017. DOI <https://doi.org/10.21165/el.v46i1.1740>

CASTILHO, A. T. O português culto falado no Brasil: história do Projeto NURC/BR. In: PRETI, D.; URBANO, H. (org.). **A linguagem falada culta na cidade de São Paulo: estudos**. vol. 4, São Paulo: FAPESP/T.A. Queiroz, 1990. p. 141-202.

DIAS, J.; CARMO, M. C. A metátese na variedade do interior paulista. **Muitas Vozes**, Ponta Grossa, v. 10, p. 1-23, 2021. DOI <https://doi.org/10.5212/MuitasVozes.v.10.2118483>

DIAS, J. R. **Que nem**: um estudo do processo de gramaticalização. 2011. 108 f. Dissertação (Mestrado em Linguística e Língua Portuguesa) – Universidade Estadual Paulista, Faculdade de Ciências e Letras, Araraquara, 2011. Disponível em: <https://repositorio.unesp.br/entities/publication/f7ce51a1-ee15-47a8-ba94-d8b7296d7a4a>. Acesso em: 18 jun. 2025.

FERREIRA, J. S. **O apagamento do /d/ em morfema de gerúndio no dialeto de São José do Rio Preto**. 2010. 142 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, 2010. Disponível em: <https://repositorio.unesp.br/entities/publication/91af3c23-0b9b-4afc-8e91-1d9ad801869c>. Acesso em: 22 abr. 2025.

FIAMENGUI, A. H. R. **A marcação de pluralidade no SN na fala e na escrita de adolescentes da região de São José do Rio Preto**. 2011. 143f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto. Disponível em: <https://repositorio.unesp.br/entities/publication/94156286-4e34-43da-b41d-7b4e01821973>. Acesso em: 22.mar. 2025.

FONSECA, A. M. H. **A perífrase verbal ir + infinitivo e o futuro do dialeto riopretano**: um estudo na interface sociolinguística/gramaticalização. 2010. 162f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista. São José do Rio Preto, 2010. Disponível em: <https://repositorio.unesp.br/entities/publication/7bad90a6-6906-4653-b3e9-0470a3ee5e08>. Acesso em: 22 mar. 2025.

FORTILLI, S. de C. **Predicados matrizes adjetivais de orações subjetivas no Português brasileiro**: gramaticalização e dessentencialização. 2013. 163f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2013. Disponível em: <https://repositorio.unesp.br/entities/publication/0532a4f7-6056-4986-9310-c80c3897f99f>. Acesso em: 22 abr. 2025.

FORTILLI, S. C. Enunciados Metadiscursivos em entrevistas do Banco de Dados Iboruna. **Signum: Estudos da Linguagem**, Londrina, n. 19/2, 2016, p. 193-214. DOI <https://doi.org/10.5433/2237-4876.2016v19n2p193>

GONÇALVES, S. C. L. **O português falado na região de São José do Rio Preto: constituição de um banco de dados anotado para o seu estudo**. Relatório científico parcial I apresentado à FAPESP. 2005. Disponível em: <http://www.alip.ibilce.unesp.br>. Acesso em: 23 jun. 2026.

GONÇALVES, S. C. L. **Banco de dados Iboruna: amostras eletrônicas do português falado no interior paulista**. Revisadas. 2007. Disponível em: <http://www.alip.ibilce.unesp.br>. Acesso em: 23 jun. 2026.

GUERRA, A. R. **Funções Textual-Interativas dos Marcadores Discursivos**. 2007. 233f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2007. Disponível em: <https://repositorio.unesp.br/entities/publication/62b07a7f-f96f-4c68-9255-55150508340e>. Acesso em: 22 abr. 2025.

KAPP-BARBOZA, A. M. M. **Usos do verbo saber e a expressão de evidencialidade no português brasileiro**. 2017. 165 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista. São José do Rio Preto, 2017. Disponível em: <https://repositorio.unesp.br/entities/publication/d6bc3900-8e3a-4bc9-be20-0ad271469003>. Acesso em: 18 jun. 2025.

LOPES-DAMÁSIO, L. R. **Diacronia dos processos constitutivos do texto relativos a *assim*: um novo enfoque da gramaticalização**. 2011. 287f. Tese (Doutorado em Estudos Linguísticos). Universidade Estadual Paulista. São José do Rio Preto, 2011. Disponível em: <https://repositorio.unesp.br/entities/publication/65341dfe-39d2-4f69-8e4e-8db7fdd915a1>. Acesso em: 22 abr. 2025.

MARCATO, F. **Análise prosódica de preposições monossilábicas**. 2013. 158 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2013. Disponível em: <https://repositorio.unesp.br/entities/publication/f55769ae-35d1-48f6-b2ff-331860711701>. Acesso em: 18 jun. 2025.

NAGAMURA, G. H. **A expressão da subjetividade na Gramática Discursivo-Funcional**. 2016. 167f. Tese (Doutorado em estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2016. Disponível em: <https://repositorio.unesp.br/entities/publication/28aac072-3409-4029-8f36-cd443043dd0e>. Acesso em: 22 jun. 2025.

PAVEZZI, V. C. **A haplologia na variedade paulista**. 2006. 116f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2006. Disponível em: <https://repositorio.unesp.br/entities/publication/31f17694-cc29-4ee9-930c-05721cfda828>. Acesso em: 20 ago. 2024.

PENHAVEL, P. **Marcadores discursivos e articulação tópica**. 2010. 168f. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, 2010. Disponível em: <https://hdl.handle.net/20.500.12733/1611844>. Acesso em: 4 jun. 2025.

RAMOS, A. P. **Descrição das vogais postônicas não-finais na variedade do noroeste paulista**. 2009. 177f. Dissertação (Mestrado em Estudos Linguísticos – Universidade Estadual Paulista, São José do Rio Preto, 2009. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/c02804ca-410e-4431-a7a1-838a7e077392/content>. Acesso em: 22 mai. 2025.

RUBIO, C. F. **A concordância verbal na região noroeste do Estado de São Paulo**. 2008. 152f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008. Disponível em: <https://repositorio.unesp.br/entities/publication/76f07ba3-7d4b-4241-8aef-9b06170110dd>. Acesso em: 15 abr. 2025.

RUBIO, C. F. **Padrões de concordância verbal e de alternância pronominal no português brasileiro e europeu: estudo sociolinguístico comparativo**. 2012. 391f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2012. Disponível em: <https://repositorio.unesp.br/entities/publication/80547951-e246-49a1-92f8-0bd2297cbfb6>. Acesso em: 15 abr. 2025.

SALOMÃO, M. H. **A variação de pluralidade nas estruturas predicativas da variedade falada na região de São José do Rio Preto**. 2010. 162f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/2a65166f-c65c-4e76-bf08-899c6c1757b6/content>. Acesso em: 22 mar. 2025.

SANTANA, L. **Motivações funcionais da gradação entre construções encaixadas nominais e verbais**. 2010. 261 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010. Disponível em: <https://repositorio.unesp.br/entities/publication/27b6cb6f-26bf-401d-8f96-cd750de1b793>. Acesso em: 22 abr. 2025.

SANTOS, R. M. A. **O uso variável do modo subjuntivo em estruturas complexas**. 2005. 144f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2005. Disponível em: <https://repositorio.unesp.br/entities/publication/3d3b7657-897c-42c1-8b13-a75db8003137>. Acesso em: 22 abr. 2025.

SILVEIRA, A. A. M. **As vogais pretônicas na fala culta do noroeste paulista**. 2008. 143 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2008. Disponível em: <https://repositorio.unesp.br/entities/publication/957a32c2-6b7c-4f46-ab28-ac3555491f61>. Acesso em: 20 mar. 2025.

VENDRAME, V. **Os verbos ver, ouvir e sentir e a expressão da evidencialidade em língua portuguesa**. 2010. 176 f. Tese (Doutorado em Estudos Linguísticos) – Universidade Estadual Paulista, São José do Rio Preto, 2010. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/d199bb4d-0426-4caa-82c8-c115b0607f0d/content>. Acesso em: 17 jun. 2025.

WIEDEMER, M. L. A correlação entre o processo de hesitação e o tipo textual: aspectos da fala do interior paulista. **Revista de Letras**, v. 21, n. 32, p. 21-44, 2019. DOI <https://doi.org/10.3895/rl.v21n32.9720>